# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-331012

(43)Date of publication of application : 30.11.2000

(51)Int.Cl. G06F 17/30

(21)Application number : 11-138070

(22)Date of filing : 19.05.1999

(71)Applicant : OKI ELECTRIC IND CO LTD

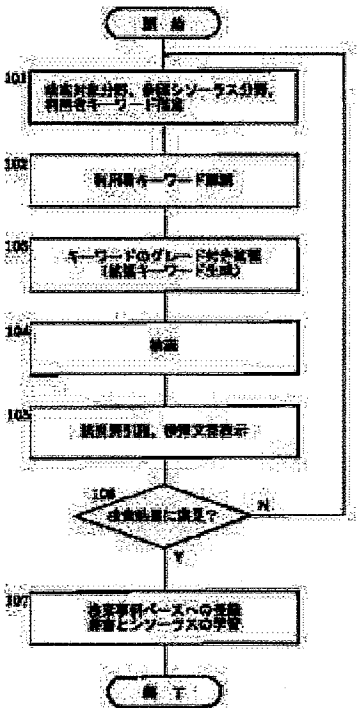
(72)Inventor : JIYOUFUU TOSHIHIKO

## (54) ELECTRONIC DOCUMENT RETRIEVAL METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To rationally and quickly retrieve a desired document.

SOLUTION: When a retrieval object classified to a field, a retrieval object in the desired field out of thesauruses, a thesaurus, and a desired keyword are designated (step 101), a word selected on the basis of the retrieval object, the thesaurus designating field, and a dictionary is logically combined with the keyword to obtain an extended keyword (step 103), and an index word in the retrieval object designating field in an index where the index word generated in accordance with the retrieval object field and information which specifies documents including this index word correspond to each other and the extended keyword are used to retrieve a pertinent document (step 104), and this document is displayed on a monitor screen together with the index word used for retrieval and the grade (step 105), and learning of the dictionary and thesauruses is performed (step 107).



* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

---

[Claim(s)]

[Claim 1]By specifying two or more retrieval objects and thesauri of the request field as a seed respectively out of a retrieval object and a thesaurus by which the field division was carried out, and specifying a desired keyword, A word selected based on each specification field and a dictionary set up beforehand of said retrieval object and a thesaurus, A search word in the specification field of said retrieval object in an index on which a search word which carried out logical combination to said specified keyword, obtained an extended keyword, and was created according to each retrieval object field, and information which specifies a document containing the search word are made to come to correspond, Search document corresponding using said extended keyword, and a searched document, It displays on monitor display with a grade called for by a search word and the computing method set up beforehand which were used for the search, And an electronic document search method performing study of said dictionary and a thesaurus based on a search word selected arbitrarily in a search word displayed on said specified keyword, said extended keyword obtained from this keyword, and said monitor display.

[Claim 2]By specifying a retrieval object of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a thesaurus and by said each specification of these retrieval objects, a keyword, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of said retrieval object and a thesaurus to said specified keyword, and obtaining an extended keyword.

[Claim 3]By specifying a thesaurus of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a retrieval object and by said each specification of these thesauri, a keyword, and a retrieval object. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of said thesaurus and a retrieval object to said specified keyword, and obtaining an extended keyword.

[Claim 4]In the electronic document search method according to claim 1, by specifying a desired keyword, It is specified by each field of a retrieval object and a thesaurus, and by said each specification of these keywords, a retrieval object, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of said retrieval object and a thesaurus to said specified keyword, and obtaining an extended keyword.

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

[Detailed Description of the Invention]
[0001]
[Field of the Invention]This invention relates to the electronic document search method which searches a desired document using a keyword out of two or more electronic documents (it is also only called a document in this specification.).
[0002]
[Description of the Prior Art]In recent years, progress of the document electronization which records on a storage or draws [ request / of paperless issue etc. ] up a document on a storage as electronic data from the start by using the contents of the paper document as electronic data is remarkable. Since the electronic document on the above-mentioned storage can also perform the search electronically, when searching the document (document which suited the user's retrieval object) desired out of a lot of documents, compared with a paper document, its usefulness is very high.
[0003]Search of an electronic document conventionally the word and the phrase (in this specification, it is only called the keyword.) used as a key, The retrieval required using what was combined with logical symbols, such as NOT which means that AND which means that two or more keywords are contained in the same document, OR which means that either of two or more keywords is contained in the same document, or a keyword is not contained in a document, is performing. Since a user's burden of keyword selection becomes large only by a keyword, there are also the existing dictionary and a thing which extends a keyword mainly using English-Japanese, a Japanese-English dictionary, or a synonym dictionary (thesaurus), and was searched. Not the binary of 1 (truth) or 0 (imitation) but the continuous value of a before [ 0-1 ] is taken to a keyword, and the fuzzy search whose search of the document containing the search word near 1 is enabled is also proposed.
[0004]conventionally the field (retrieval object field) to search is not divided into some, and it searches to one field common about any keywords, i.e., one huge field, and a thesaurus is also in the situation where it is only that there is what [ a ] is common to all the fields.

[0005]
[Problem(s) to be Solved by the Invention]Since the new word, technical term, or foreign language which is not contained in an existing dictionary or thesaurus is not considered at all by conventional technology as mentioned above, Search of the document which cannot respond to a huge new word, a technical term, or a foreign language even if it uses the method of extending and searching a keyword, using the dictionary and thesaurus of these existing, but contains a new word etc. was impossible or remarkably difficult.
[0006]Since it would always search for [ all the ] a field even when the retrieval object field of the document to desire can predict to some extent, since there was only a thing which has the retrieval object field and a thesaurus common to all the fields, search took time.
[0007]Use of only mere English-Japanese and a Japanese-English dictionary is insufficient about extension of a keyword, therefore, pass use of the reading (Chinese character kana) dictionary of a Chinese character, or an English reading (English kana) dictionary, or proofreading [ as / in WWW or Net News ] -- use of many dictionaries, such as use etc. of the misspelling dictionary to the document which is not, can be considered. If not only the dictionary that extracts such synonymous words but the various thesauri which extract a synonym will be used, the total of the dictionary used for search or a thesaurus will increase. Then, the user needed to choose which dictionary and thesaurus are used and required time and effort for the preparation before retrieval execution. The efforts of search did not bear fruit considering time and effort, and the document to desire was not obtained.
[0008]This invention is made that the problem of the above-mentioned conventional technology should be solved.
[0009]
[Means for Solving the Problem]The next composition is used for this invention in order to solve above-mentioned SUBJECT.
<Composition 1> by specifying two or more retrieval objects and thesauri of the request field as a seed respectively out of a retrieval object and a thesaurus by which the field division was carried out, and specifying a desired keyword, A word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus, A search word in the specification field of the above-mentioned retrieval object in an index on which a search word which carried out logical combination to the specified above-mentioned keyword, obtained an extended keyword, and was created according to each retrieval object field, and information which specifies a document containing the search word are made to come to correspond, Search document corresponding using the above-mentioned extended keyword, and a searched document, It displays on monitor display with a grade called for by a search word and the computing method set up beforehand which were used for the search, And an electronic document search method performing study of the above-mentioned dictionary and a thesaurus based on a search word selected arbitrarily in a search word displayed on the specified above-mentioned keyword, the above-mentioned extended

keyword obtained from this keyword, and the above-mentioned monitor display.
[0010]<Composition 2> by specifying a retrieval object of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a thesaurus and by each above-mentioned specification of these retrieval objects, a keyword, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus to the specified above-mentioned keyword, and obtaining an extended keyword.
[0011]<Composition 3> by specifying a thesaurus of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a retrieval object and by each above-mentioned specification of these thesauri, a keyword, and a retrieval object. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned thesaurus and a retrieval object to the specified above-mentioned keyword, and obtaining an extended keyword.
[0012]<Composition 4> in the electronic document search method according to claim 1, by specifying a desired keyword, It is specified by each field of a retrieval object and a thesaurus, and by each above-mentioned specification of these keywords, a retrieval object, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus to the specified above-mentioned keyword, and obtaining an extended keyword.
[0013]
[Embodiment of the Invention]Hereafter, it explains using a drawing per example of this invention.
<<Example 1>>
<The composition of the example 1 and operation> The flow chart which shows the example 1 of the electronic document search method according [ drawing 1 ] to this invention, and drawing 2 are the explanatory views of a search system in which the example 1 of this invention method was applied. As shown in drawing 2, a search system here is provided with the keyword interpretation part 22, the thesaurus 23 (23a - 23c--), the dictionary 24, the search case base 25, the retrieval part 26, and the index 27.
[0014]The above-mentioned keyword interpretation part 22 is a formation part which receives the retrieval required which consists of the keyword for search specified by the user 21 (user keyword), a retrieval object field, and a field (reference thesaurus field) of the thesaurus 23 to refer to, and gives an extended keyword and the retrieval object field to the retrieval part 26. Logical combination of two or more keywords is carried out, and a user keyword usually contains the sign (wild card) of match partial here. The keyword interpretation part 22 interprets the specified user keyword and the user keyword which contained the sign (wild

card) of match partial here, It has the function to extract the pure keyword which separated the function and wild card which distinguish the kinds (full match, prefix search, etc.) of match partial. The keyword interpretation part 22 checks whether the group of the extracted keyword exists in the search case base 25 as it is, If it exists, the group of the keyword (extended keyword) after the extension in the example will be extracted, it is shown to the user 21 via monitor display (not shown), and it has a function which waits for correction if needed for the user 21, and is given to the retrieval part 26. If the group of the above-mentioned keyword does not exist in the search case base 25 as it is, It is checked respectively whether those keywords are registered into the thesaurus 23 or the dictionary 24 specified by the user 21 as an entry, If registered, it will give the retrieval part 26 as an extended keyword by which OR combination was carried out with the word (synonymous words, synonym) acquired from the entry, and each of the keyword which makes the above-mentioned group.

[0015]the above-mentioned thesaurus 23 is beforehand created for every retrieval object field here like the computer field thesaurus 23a, the scientific discipline thesaurus 23b, and the social field thesaurus 23c, and is a fuzzy thesaurus with the similarity of the synonym to an entry. The above-mentioned dictionary 24 is provided with English-Japanese, Japanese-English, Chinese character kana, English kana, a misspelling, and six dictionaries of an abbreviation here.

[0016]The above-mentioned search case base 25 the past example of search (search example) 1 or plurality, Here, it has memorized in the group of two or more keywords, the retrieval object field, and the reference thesaurus field, among those if even two are specified by the user 21, it also has the function to specify the one [ remaining ] automatically, and it is used for the complement of retrieval required. for example, the search case base 25 -- keyword:homepage creation retrieval object field: -- network reference thesaurus field:, when there is an example of search of the past called science, Keyword: Homepage Creation retrieval-object field: If retrieval required is specified to a network, reference thesaurus field:science will be complemented and it will display on monitor display, and if the user's 21 check can be taken, automatic setting of the "science" will be carried out as a reference thesaurus field.

[0017]The retrieval part 26 is a formation part which receives the extended keyword and the retrieval object field from the keyword interpretation part 22, and searches document corresponding with reference to the index 27. Namely, the inside of the index 27 with which the retrieval part 26 comprised a search word group extracted from each document in a retrieval object document group, Search results are given to the user 21 by extracting with a grade the document which measures each and the above-mentioned extended keyword of a search word in the search word group belonging to retrieval object part Nouchi's retrieval object document group given from the above-mentioned keyword interpretation part 22, and contains a search word in agreement, and displaying on monitor display. The above-mentioned grade sets to 1 about the document which fills the logical combination of the user keyword itself, for

example, and if the document by which the document containing a synonym fulfills 1 or less and a neighboring strip affair does not contain 1 or more and a keyword directly, it is made or less into one. Although the search word which was in agreement with the keyword by search is displayed on monitor display with a retrieval sentence document, If the user 21 chooses a desired search word on monitor display before long, if it is a synonym, the thesaurus of the above-mentioned retrieval object field in the thesaurus 23 will register with the dictionary 24 respectively, if it is synonymous words (study). The group of the specified keyword, the retrieval object field, and the reference thesaurus field is registered into the search case base 25 as a search example. These study and registration processing are performed by the above-mentioned keyword interpretation part 22.

[0018]The above-mentioned index 27 is created as follows here. In order to make the size small and to speed up search in creation of the index 27, it is made as [ be / no duplication of the word from which types of letters, such as a Chinese character, a hiragana, katakana, an alphabetic character, and a number, differ ], and is considered as a pause of a search word by pause of a type of letters. For example, from "information filtering", if the usual search word logging tool is used, although three, "information filtering", "information", and "filtering", will be extracted, now, the number of search words will increase, the size of the index 27 will increase, and search will come to take time. Therefore, let only two, "information" and "filtering", be a search word here. It compensates with carrying out specification (adjoining AND=NAND specification) that "information" and "filtering" are next to one word at the time of retrieval required, to the retrieval required "information filtering", and searching. Two, what sorted the search word in numerical order as it was, for example, and the thing which made the "computer" the reverse order like "machine ****", and sorted it in numerical order, are created, and the index 27 is used. Although the latter thing is used for a suffix search, it is not indispensable. The index structure for searching a document name from the search word in the index 27 is the same as that of what is used for the search in a publicly known database.

[0019]An example of the monitor display display information at the time of the retrieval-required input of the above-mentioned search system is shown in drawing 3. In this figure, white round"O" and a black dot "-" are the button switches for a selection of function respectively operated with a pointing device (not shown), a white round head expresses OFF and a black dot expresses one. here -- the 1st keyword -- "world wide web" -- the 2nd keyword -- "hp" -- AND -- it is inputted with logical combination. "Science" is inputted into the retrieval object field and "economy" is respectively inputted into the reference thesaurus field. If the search button switch B1 is operated with the above-mentioned pointing device, search will be started, if cancel button switch B-2 is operated, the operation will be canceled during all the alter operation and operations, and a keyword, a field, and a button switch will be returned to an initial state (a blank or OFF).

[0020]Hereafter, drawing 1 and drawing 3 are used together and the example 1 of this invention method is described. In Step 101, a keyword (user keyword), the retrieval object

field, and the reference thesaurus field are specified by the user 21 (input). When drawing 3 is taken and stated to an example, a user keyword, It is specified that the abbreviation "www" is extended by OR combination about the 1st user keyword (keyword 1) "world wide web" so that it may understand from the button switch for a selection of function for "carrying out OR combination of the abbreviation" being turned on ("-").

[0021]Here, it is interpreted as the user keyword which put two or more words in order and was written in 1 field being combined by AND of contiguity of each word, and, in the usual AND combination, it is supposed that the field will be changed and inputted. Therefore, fundamentally, although the 1st user keyword (keyword 1) "world wide web" illustrated to drawing 3 is interpreted as the retrieval required "search the document which exists in the same document with these three words, and appears continuously in this order", It is simultaneously interpreted as "world", "wide", "wide", and "web" being adjoining AND combination respectively. A search by such logical combination is performed according to the word order of the usual reading. The 2nd user keyword (keyword 2) "hp" is interpreted as AND combination being carried out with the above-mentioned user keyword "world wide web" which OR combination of the "www" was carried out and was extended.

[0022]Although interpretation of such a user keyword and extension will be performed at the below-mentioned step 102,103, a user keyword is specified on the monitor display shown in drawing 3 on the assumption that such an interpretation and extension are made. The above-mentioned automatic setting of a keyword, the retrieval object field, or the reference thesaurus field based on the past example of search by the search case base 25 is also performed.

[0023]Although not illustrated by drawing 3, some signs for specifying match partial etc. are usually attached, and a user keyword is inputted. As a sign (wild card) which is in agreement with the character string of (a) arbitration, for example, "*", (b). [ whether two or more words are connected with "-" into one keyword, and ] Two words exist in both ends in "_" and a (c)1 ** keyword as a sign which shows that it adjoins, or as for the 1st word, it is a backward match, "+" ** is set up as a sign which shows that these two adjoin, and prefix search and the 2nd word are suitably given to a keyword.

[0024]According to the above (a), flexible specification of prefix search, a backward match, a middle match, both-ends coincidence, etc. can be performed by "*" as well as the regular expression of UNIX shell. According to (b), effective search can be respectively performed according to the idioms (angled-shot etc.) of English connected with "-", and (c) about the idioms (the certifying examination of information processing = information processing certifying examination etc.) of Japanese connected with the particle. For example, a keyword with a word "information" and "an examination" is set to information * examination OR (information *NAND* examination). "Information processing examination", "the certifying examination of information", "the certifying examination of information processing", an information processing certifying examination", etc. can search simultaneously by this.

[0025]Here, it supposes that the logical combination of a user keyword is allowed to two steps,

and the 1st step of logical combination makes three kinds of AND, and adjoining AND and OR, and the 2nd step of logical combination four kinds of AND, OR, and OR and NOT of an abbreviation. For example, it becomes about NAND like the AND(world NAND wide NAND web) (home page*) AND Japan Federation of Bar Associations considering not NAND of logical symbols but adjoining AND, and AOR as OR of an abbreviation. AOR is what is used when making registration (refer to the below-mentioned step 107) to the abbreviation dictionary in the dictionary 24 into a key objective, For example, if keyword specification is carried out with AOR www (world NAND wide NAND web), "www" will certainly be registered into an abbreviation dictionary as an abbreviation of "world wide web." Here, it is supposed that OR combination will not be carried out even if "www" has other official names (word).

[0026]A user keyword is interpreted in Step 102. In addition to the interpretation of the above-mentioned user keyword, in this step 102, extraction of the kind of coincidence and a pure keyword is performed from the specified user keyword. As an example, if a user keyword is info*filter, it is a kind:both-ends coincidence keyword of coincidence. : (info, filter)
It interprets and is saved in the memory in a search system.

[0027]In Step 103, a user keyword is extended using the dictionary 24 and the specified thesaurus 23, and is given to the retrieval part 26 as an extended keyword. A user keyword is compared and contrasted with each entry in the dictionary 24 one by one, is extended by the word and OR combination which are obtained from the congruous entries (OR extension), and, specifically, is given to the retrieval part 26 as an extended keyword.

[0028]Since there are what changes a meaning by neither a field nor the context depending on the kind of dictionary 24, for example, English-Japanese, a Japanese-English dictionary, etc. and a changing thing, it is chosen which meaning the keyword interpretation part 22 takes based on the thesaurus 23 of the field specified by the user 21. For example, since it will be satisfactory for extending a user keyword "monkey" with a "monkey", OR combination of both is carried out in this case, namely, it gives the retrieval part 26 by making "monkey"OR "monkey" into an extended keyword. supposing a user keyword is "comp", since this is an abbreviation of a formal word "computer" in the field of information, it will carry out OR combination of both, namely, it will carry out OR extension with "comp"OR"computer", and will be given to the retrieval part 26 by making this into an extended keyword. If it specifies by "comp*" and prefix search, since a document including an unrelated word will be searched, "compare" etc. will not perform such extension here.

[0029]Since ambiguity is high, the abbreviation of a compound is extended as follows, for example. That is, supposing a user keyword is "hp", NAND (adjoining AND) combination will be carried out with the formal word (compound) "home page", and it will extend with "home"NAND"page." A keyword "hp" has a meaning of "home party." In this case, if the keyword interpretation parts 22 are a computer and a network as it is "home party" if the retrieval object fields specified by the user 21 are amusement and a life, they will distinguish that it is "home page", carry out NAND combination like the above, and are extended.

[0030]A grade is given to the word by which OR combination is carried out. Although this grade is called for by calculation, the thing used as that antecedent basis is the connection probability of a word here. "Connection" means that the words, for example, a word, "i", and the word "j" approach and appear, and the word "j" can consider various modes to be the words "i" like an appearance (coincidence) in an appearance (coincidence) or a specific document set in an appearance and 1 document within what [ order ] word. a mode with proper connection probability, and here -- contiguity (it appears in one word approximately) -- a mode is chosen, and it asks by a lower formula (1), and is used as the above-mentioned grade.
Connection probability Wij=(number of times which word [ "i" ], and word "j" connected)/ (number of times to which one of the word "i", and the words "j" appeared) -- (1)
Such connection probability (grade) is called for for every field.
[0031]The connection probability Wij of a certain keyword ki and a certain word "kj" is expressed like "electronic" 0.5 "network" 0 and 3 "reception" 0.2, for example to a keyword "e-mail."
[0032]calculating and saving connection probability about all the appearing words -- therefore, if it takes [ that the memory space to need becomes great or ] into consideration how much the obtained connection probability becomes what can set reliance practically, it will not necessarily be hard to say a best policy. Therefore, grouping of two or more words of the actual almost same meaning is carried out, connection probability is calculated about one word (representation word) of them, and it is considered as the connection probability of all the words which belong the value to a group. For example, a "network group" is formed by a "network", "Network", "****", etc., The connection probability called for about the above "network" before long is used also as connection probability, such as "other words in a "network group", i.e., "Network", and ****." These "network", "Network", "****", etc. are registered with connection probability (grade) into the dictionary 24 as synonymous words here.
[0033]The "computer" should be specified as a user keyword now, and in the dictionary 24. Computer and a computer due to English versus Japanese due to English versus katakana, In Computer and a computer, a computer and **** Mr. ** assume that the electronic computer and the computer were registered as a Japanese abbreviation, and the computer and the computer were respectively registered for Computer and Comp. as fluctuation of the notation as an English abbreviation due to Japanese versus a hiragana. On the other hand, since a "network" and "Network" are registered into the dictionary 24 as synonymous words as mentioned above, "Computer Network" and a "computer network" can calculate the connection probability (grade) as the same thing. By this, "Computer Network" and a "computer network" will be searched with the same grade. Namely, a "computer network" will be searched if "ComputerNetwork" is searched, When a grade is given to search results (document) so that it may mention later, the grade of the value with same document containing "Computer Network" and document containing a "computer network" will be attached.

[0034]Next, a user keyword "hp" is explained. "hp" shall be an abbreviation and there shall be two, "home page" and "home party", as a formal word (compound). Now, in the computer field thesaurus 23a, supposing it appears 100 times as "home party" 300 times as "home page", if the grade in the computer field of each word in that case considers it, for example as grade = (appearance frequency of word to search for)/(the maximum appearance frequency of one of words) -- "home page" -- it becomes :300/300=1"home party":100/300=0.333 --. therefore, the document containing "home page" will be searched with a certainly bigger grade than the document containing "home party" by specifying the computer field thesaurus 23a. When a grade is given to search results (document), a grade will be attached by the ratio according to the size of the above-mentioned grade of the document containing "home page" and the document containing "home party."

[0035]it may be better to use the thesaurus 23 of a different field from the retrieval object field on search of the document which the direction which the above-mentioned example is what described the case where it was the compound (idiom) which the word adjoined, and used the thesaurus 23 of the retrieval object field and a field in agreement in this case desires, although it shall be effective It is obvious in the computer field that a personal computer (personal computer) is machinery, therefore, the thing done for OR extension of the synonym "machinery" from a user keyword "personal computer" using the thesaurus 23 of the computer field it and whose retrieval object field correspond -- it is because it is thought that things are difficult. In such a case, in order to change a viewpoint, OR extension of the synonym is carried out using the thesaurus 23 of a different field from the retrieval object field which is called the economic-categories thesaurus and amusement field thesaurus (neither is illustrated) in the thesaurus 23, for example. when it is the words which have the hierarchical order in a concept like a "personal computer" and "machinery", such OR extension is effective also from the point that search results of the direction which carried out OR extension of the synonym using the thesaurus 23 of a different field from the retrieval object field improve in many cases -- it can say. Therefore in this example 1, are carrying out [ that it is selectable (refer to Step 101) and ] arbitrarily the thesaurus field referred to in retrieval required. In OR extension of the synonym in the word of the level without the hierarchical order of a concept like a "personal computer" and a "workstation", the basic technique of coinciding the retrieval object field and the reference thesaurus field is protected.

[0036]The connection probability Wij which can be found by an upper formula (1) also expresses the similarity of a certain word and other words. Therefore, the upper equation (1) which is a formula of the connection probability Wij is applicable also to the similarity calculation of the synonym in the thesaurus 23.

[0037]A document retrieval is performed in Step 104. The extended keyword from the keyword interpretation part 26 is compared and contrasted with each search word in the index 27 one by one, and, specifically, extraction of the document containing the congruous search words is performed. The grade which an extended keyword has is given to the extracted document as a

grade of the document. Publicly known heuristics, for example, binary search, is used for search.

[0038]Supposing "info*" is contained in the extended keyword now, this is a prefix search of "info", and it is infoinfo in this case, for example.

It becomes a search word to which informinfomationinformation corresponds, The extraction result (the number of document extraction to an applicable search word) of info 500 document info.200 document inform 300 document infomation100 document information 1056 document etc. is obtained.

[0039]And to the document containing each search word, by the logical combination of other keywords in an extended keyword, it calculates further (it is a product set etc. at a set union and AND in OR), and the document group to desire is narrowed down and extracted. As for AND, an arithmetic product, the minimum, and OR calculate an arithmetic sum, the maximum, and NOT with a difference. It searches by assuming it first to be the usual AND (it does not adjoin), and in the case of adjoining AND, it confirms whether two keywords connected with AND into the document group extracted by this next adjoin, it extracts an adjoining document, and makes it search results. Although these search results are expressed to monitor display as Step 105, it searches assuming that it is the usual AND, and may be made to display the extracted document group as search results. Under the present circumstances, if the grade of the adjoining document is raised, distinction with the document which does not adjoin will become easy. Here, the grade of the adjoining document is increased 1.1 times and facilities are given by the display of the search results according to the stage of search, and the check.

[0040]The display of search results is performed in Step 105. That is, an end of search will display the document (retrieval sentence document) extracted by that cause on monitor display. In order that the retrieval sentence document which is search results may tell whether it is what was searched with what kind of search word, a retrieval sentence document is displayed by correspondence with the search word used for the search. The grade is attached and each retrieval sentence document is displayed. Let the grade here be a value required in carrying out the multiplication of each above-mentioned grade (value) calculated in the extraction process of each retrieval sentence document here [, such as addition or multiplication, ].

[0041]The display of the number [ first as opposed to each search word in the display of a retrieval sentence document ] of retrieval sentence documents, the display of the document name extracted by the search word by directing a specific search word to the next with a pointing device (not shown), Then, various stage displays, such as a display etc. of the applicable page (page the search word is described to be) of the document by directing a specific document name with the above-mentioned pointing device, are possible. The information which can specify the document eventually should just be displayed as a display of a retrieval sentence document, For example, in addition to these titles, if it is books and the title of books, an author, a date issued, a publishing office, etc. are magazines, if a serial

number is a paper, an academic journal name, a title of a paper, a presenter, a date issued, a publishing office, etc. where it appeared correspond.

[0042]In Step 106, the check of search results (retrieval sentence document) is performed. That is, in this step 106, the judgment of whether to have been satisfied with search results is made. Search results are expressed as Step 105, and when the document to desire is obtained, processing moves to Step 107 by operation of the purport are satisfied with the user's 21 search results. When the user 21 is not satisfied with search results, Steps 101-106 are repeated -- the document to desire is not obtained -- until it is satisfied. Since the user 21 is to leave the keyword (user keyword), retrieval object field name, and reference thesaurus field name which were specified first here to the buffer memory of a search system, It ends with fine adjustment of carrying out additional specification of the keyword, or usually carrying out change specification of the above-mentioned each field name as operation of the user 21 at the time of a search repetition.

[0043]In Step 107, study of the dictionary 24 and the thesaurus 23 and case registration to the search case base 25 are performed. Specifically, the abbreviation by which OR combination was carried out is registered into a user keyword as an abbreviation of the user keyword before combination by the abbreviation dictionary in the dictionary 24. For example, the place which searched by carrying out OR combination of the abbreviation [ keyword / user / "world wide web" ] "www", When the document to desire is extracted, "www" is registered into the abbreviation dictionary in the dictionary 24 as an abbreviation of "world wide web" (when judged with result satisfaction in Step 106). "www" is registered into the abbreviation dictionary in the dictionary 24 from the start as an abbreviation of "world wide web", and drawing 3 is what illustrated the case carried out so that OR combination of the abbreviation "www" may be carried out by one of the button switch for a selection of function, and differs from an example here.

[0044]Although the search word which was in agreement with the keyword by search is displayed on monitor display in Step 105, If the user 21 chooses a desired search word on monitor display before long, if it is a synonym, the thesaurus of the above-mentioned retrieval object field in the thesaurus 23 will register with the dictionary 24 respectively, if it is synonymous words (study).

[0045]The thesaurus 23 specified by the user 21 is also learned as follows by the contents of the keyword (user keyword) which the user 21 specified. That is, in this case, although many user keywords are given from many user 21 -- to the search system concerned after operation of a search system, it enlarges connection probability noting that both the keywords of NAND (adjoining AND) joint specification have high similarity. For example, when user keywords are "multi-"NAND" media", new connection probability Wij' as Ki= multi and Kj= media, Wij'=wij+ {(number of times in which Ki and Kj carried out AND combination) /(number of times to which one of Ki and the Kj(s) appeared)} x (1-Wij) -- Formula (4)

It carries out. Now, to keyword" multi-", supposing connection probability Wij' of a keyword

"media" is 0.75, it will register with the thesaurus 23 noting that the similarity of the "media" to "multi-" or "media" or "multi-" is 0.75 (study). The similarity between the keywords which this connects well becomes large, and at the time of subsequent search. When either one of "multi-" or "media" becomes a user keyword or it is contained in a user keyword, the grade of the document containing the word "media" of the other or "multi-" becomes large, and it becomes easy to carry out search as a document which the document desires.

[0046]The group of the specified keyword, the retrieval object field, and the reference thesaurus field is registered into the search case base 25 as a search example. These study and registration processing are performed by the above-mentioned keyword interpretation part 22.

[0047]<Effect of the example 1> As opposed to the retrieval required [ according to / as stated above / the example 1 ] by specification of the retrieval object field, the reference thesaurus field, and a user keyword, Since it was made to display on monitor display with the search word and grade which used the searched document for the search, it is effective in the ability of search of only the document to desire to carry out more appropriately than the conventional method. Since the dictionary 24 and the thesaurus 23 which were used learn automatically based on the search word selected arbitrarily in the search word displayed on a user keyword, an extended keyword, and monitor display etc., Search of the document which the know how of retrieval methods, such as the user 21 keyword specification and extension, could be accumulated, and search of the document to desire improved, and contained the new word, the technical term, and the foreign language is effective in the ability to carry out now more flexibly than before and easily. If it specifies sequentially from the large reference thesaurus field (or the retrieval object field) of similarity with the retrieval object field (or the reference thesaurus field) and search is advanced, it is effective in search time shortening all the fields rather than a system conventionally which is searched at once.

[0048]<<Example 2>>

<The composition of the example 2 and operation> The explanatory view of a search system in which, as for the flow chart which shows the example 2 of the electronic document search method according [ drawing 4 ] to this invention, and drawing 5, the example 2 of this invention method was applied, and drawing 6 are the figures showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 5. In these drawing 4 - drawing 6, identical codes are given to drawing 1 - a portion the same as that of drawing 3, or considerable, and the explanation is omitted. As shown in drawing 5, a search system here is provided with the keyword interpretation part and reference thesaurus field specification part 51, the thesaurus 23 (23a - 23c--), the dictionary 24, the search case base 25, the retrieval part 26, and the index 27.

[0049]The above-mentioned keyword interpretation part and reference thesaurus field specification part 51 is a formation part which receives the retrieval required which consists of the user keyword and the retrieval object field specified by the user 21, chooses and specifies

the reference thesaurus field, and gives an extended keyword and the retrieval object field to the retrieval part 26. That is, in the example 2, automatic setting of the reference thesaurus field is carried out with the keyword interpretation part and reference thesaurus field specification part 51, without the user 21 specifying (refer to Step 401,402 in drawing 4). Therefore, unlike drawing 3 in drawing 6, there is no specification display window of the reference thesaurus field in the right-hand of the specification display window of the retrieval object field. The keyword interpretation part and reference thesaurus field specification part 51 is constituted with the same function as the keyword interpretation part 22 of the example 1 besides the automatic setting function of the above-mentioned reference thesaurus field. Other portions are the same as that of drawing 2 among drawing 5.

[0050]Below, the automatic setting of the reference thesaurus field by the keyword interpretation part and reference thesaurus field specification part 51 is explained. First, since the retrieval object field is specified by the user 21, the thesaurus of the same field as the retrieval object field is the 1st candidate of the thesaurus which carries out automatic setting with the above-mentioned reference thesaurus field specification part 51. It is mentioned as a candidate also with a strong thesaurus of a useful field for search. It is thought that it can divide roughly into two kinds, the thesaurus of a field which presents the synonym connected with search in associative memory as a thesaurus of a useful field, and the thesaurus of a field which presents a different synonym from a viewpoint or a conceptual level. the similarity between the retrieval object fields is calculated and, as for the thing of similarity which has the large former, and the latter, the small thing of similarity corresponds.

[0051]The similarity between fields can be found, for example with the following calculation methods. First, a field is vectorized. Only the number in many order extracts the word which appears in each field, and it normalizes. However, particles which come out frequently except. Here, suppose that five words are extracted in many order, and let these words be basic words. For example, the appearance frequency in the network field is e-mail 3 system 2isdn 2internet 1 cellular-phone 1, and the vector of the computer field presupposes that it is scsi 4 file 2 soft 1 system 1isdn 1.

[0052]Next, the similarity between fields is calculated. Similarity between two fields = it is considered as the inner product of an item which corresponded. Although a conflicting item is used for absolute value calculation of normalization of a vector, it is not used for the molecule of an inner product. Words tautological in a biparite field in the above-mentioned example are "isdn" and a "system", and it is in the network field (2, 2).

In the computer field (1, 1)

It is ********. Therefore, similarity is set to (2x1+2x1)/(square root of square root x23 of 19) =0.19 as x=3x3+2x2+2x2+1x1+1x1=19y=4x4+2x2+1x1+1x1+1x1=23. It asks for similarity by the vector of 1000 to 5000 word actually.

[0053]if it states concretely, it will be thought that use of the thesaurus of the large field of similarity with the retrieval object field is useful when obtaining a detailed additional keyword,

and use of the thesaurus of the small field of similarity is useful to conversion of a viewpoint. For example, although a "terminal" means a "computer" to the user 21 of computer business, it means a "telephone" to the user 21 of a telephone industry. Since it becomes difficult to search the document from a different viewpoint so that it is a special field, if the retrieval object field specified by the user 21 is a special field, as for a certain forge fire, the above-mentioned reference thesaurus field specification part 51 will be set up so that the thesaurus of the small field of similarity may be chosen and specified. on the contrary, the retrieval object field specified by the user 21 -- **** -- when it is a general field, the above-mentioned reference thesaurus field specification part 51 will be set up so that the thesaurus of the same or large field of similarity as the retrieval object field may be chosen and specified.

[0054]The keyword interpretation part and reference thesaurus field specification part 51, A user keyword is extended using the thesaurus 23 of the field specified by the dictionary 24 and self, For example, after displaying some extended keywords on monitor display and making the user 21 choose, The retrieval part 26 is made to refer to the procedure same about retrieval object part Nouchi's retrieval object document group specified by the user 21 as the example 1, and, finally study of the dictionary 24 and the thesaurus 23 and case registration to the search case base 25 are performed.

[0055]<Effect of the example 2> Since the automatic setting of the reference thesaurus field was made to be carried out by specifying the retrieval object field and a user keyword according to the example 2 as stated above, By setting up suitably the selection technique of the reference thesaurus field by which automatic setting is carried out, it is effective in the ability to carry out without the search which converted the detailed search by the thesaurus of the large field of similarity or the viewpoint by the thesaurus of the small field of similarity applying the time and effort of the user's 21 reference thesaurus field specification. In addition, there is the same effect as the example 1. If it specifies sequentially from the large reference thesaurus field of similarity with the retrieval object field and search is advanced, it is effective in search time shortening all the fields rather than a system conventionally which is searched at once.

[0056]<<Example 3>>

<The composition of the example 3 and operation> The explanatory view of a search system in which, as for the flow chart which shows the example 3 of the electronic document search method according [ drawing 7 ] to this invention, and drawing 8, the example 3 of this invention method was applied, and drawing 9 are the figures showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 8. In these drawing 7 - drawing 9, identical codes are given to drawing 1 - a portion the same as that of drawing 3, or considerable, and the explanation is omitted. As shown in drawing 8, a search system here is provided with the keyword interpretation part and retrieval object field specification part 81, the thesaurus 23 (23a - 23c--), the dictionary 24, the search case base 25, the retrieval part 26, and the index 27.

[0057]The above-mentioned keyword interpretation part and retrieval object field specification part 81 is a formation part which receives the retrieval required which consists of the user keyword and the reference thesaurus field specified by the user 21, chooses and specifies the retrieval object field, and gives an extended keyword and the retrieval object field to the retrieval part 26. That is, in the example 3, automatic setting of the retrieval object field is carried out with the keyword interpretation part and retrieval object field specification part 81, without the user 21 specifying (refer to Step 701,702 in drawing 7). Therefore, unlike drawing 3, there is no specification display window of the retrieval object field in drawing 9. The keyword interpretation part and retrieval object field specification part 81 is constituted with the same function as the keyword interpretation part 22 of the example 1 besides the automatic setting function of the above-mentioned retrieval object field. Other portions are the same as that of drawing 2 among drawing 8.

[0058]Below, the automatic setting of the retrieval object field by the keyword interpretation part and retrieval object field specification part 81 is explained. First, since the reference thesaurus field is specified by the user 21, the same retrieval object field as the reference thesaurus field is the 1st candidate of the retrieval object field which carries out automatic setting with the above-mentioned retrieval object field specification part 81. It is mentioned as a candidate also with the useful strong retrieval object field for search. It is thought that it can divide roughly into two kinds, the retrieval object field which presents the synonym connected with search in associative memory as a useful retrieval object field, and the retrieval object field which presents a different synonym from a viewpoint or a conceptual level. The similarity between the reference thesaurus fields is calculated and, as for the thing of similarity which has the large former, and the latter, the small thing of similarity corresponds.

[0059]The similarity between fields can be found, for example with the following calculation methods. First, a field is vectorized. Only the number in many order extracts the word which appears in each field, and it normalizes. However, particles which come out frequently except. Here, suppose that five words are extracted in many order, and let these words be basic words. For example, the appearance frequency in the network field is e-mail 3 system 2isdn 2internet 1 cellular-phone 1, and the vector of the computer field presupposes that it is scsi4 file 2 soft 1 system 1isdn 1.

[0060]Next, the similarity between fields is calculated. Similarity between two fields = it is considered as the inner product of an item which corresponded. Although a conflicting item is used for absolute value calculation of normalization of a vector, it is not used for the molecule of an inner product. Words tautological in a biparite field in the above-mentioned example are "isdn" and a "system", and it is in the network field (2, 2).

In the computer field (1, 1)

It is ********. Therefore, similarity is set to (2x1+2x1)/(square root of square root x23 of 19) =0.19 as x=3x3+2x2+2x2+1x1+1x1=19y=4x4+2x2+1x1+1x1+1x1=23. It asks for similarity by the vector of 1000 to 5000 word actually.

[0061]if it states concretely, it will be thought that the specification of the large retrieval object field of similarity with the reference thesaurus field is useful when obtaining a detailed additional keyword, and specification of the small retrieval object field of similarity is useful to conversion of a viewpoint. For example, although a "terminal" means a "computer" to the user 21 of computer business, it means a "telephone" to the user 21 of a telephone industry. Since it becomes difficult to search the document from a different viewpoint so that it is a special field, if the reference thesaurus field specified by the user 21 is a special field, as for a certain forge fire, the above-mentioned retrieval object field specification part 81 will be set up so that the small retrieval object field of similarity may be chosen and specified. on the contrary, the reference thesaurus field specified by the user 21 -- **** -- when it is a general field, the above-mentioned retrieval object field specification part 81 will be set up so that the same or large retrieval object field of similarity as the reference thesaurus field may be chosen and specified.
[0062]First the above-mentioned retrieval object field specification part 81 the largest retrieval object field of similarity, then, the next -- three sorts of retrieval object fields with large similarity -- choose and make it specified that it says from descending of similarity, or the smallest retrieval object field of similarity chooses it as the beginning or the last, and is specified as it, or various setting out is possible. The keyword interpretation part and retrieval object field specification part 81, A user keyword is extended using the thesaurus 23 of the field specified by the dictionary 24 and the user 21, For example, after displaying some extended keywords on monitor display and making the user 21 choose, The retrieval part 26 is made to refer to the procedure same about retrieval object part Nouchi's retrieval object document group specified by self as the example 1, and, finally study of the dictionary 24 and the thesaurus 23 and case registration to the search case base 25 are performed.
[0063]<Effect of the example 3> Since the automatic setting of the retrieval object field was made to be carried out by specifying the reference thesaurus field and a user keyword according to the example 3 as stated above, By setting up suitably the selection technique of the retrieval object field by which automatic setting is carried out, it is effective in the ability to carry out without the search which converted the detailed search by the large retrieval object field of similarity or the viewpoint by the small retrieval object field of similarity applying the time and effort of the user's 21 retrieval object field specification. In addition, there is the same effect as the example 1. If it specifies sequentially from the large retrieval object field of similarity with the reference thesaurus field and search is advanced, it is effective in search time shortening all the fields rather than a system conventionally which is searched at once.
[0064]<<Example 4>>
<The composition of the example 4 and operation> The flow chart which shows the example 4 of the electronic document search method according [ drawing 10 ] to this invention, The explanatory view of a search system in which, as for drawing 11, the example 4 of this invention method was applied, and drawing 12 are the figures showing an example of the monitor display display information at the time of the retrieval-required input of the search

system shown in drawing 11. In these drawing 10 - drawing 12, identical codes are given to drawing 1 - a portion the same as that of drawing 3, or considerable, and the explanation is omitted. As shown in drawing 11, a search system here is provided with the keyword interpretation part and retrieval object field, the reference thesaurus field specification part 100, the thesaurus 23 (23a - 23c--), the dictionary 24, the search case base 25, the retrieval part 26, and the index 27.

[0065]The above-mentioned keyword interpretation part and retrieval object field and the reference thesaurus field specification part 100, It is a formation part which receives the retrieval required which consists of a user keyword specified by the user 21, chooses and specifies the retrieval object field and the reference thesaurus field, and gives an extended keyword and the retrieval object field to the retrieval part 26. That is, in the example 4, automatic setting of the retrieval object field and the reference thesaurus field is carried out with the keyword interpretation part and retrieval object field and the reference thesaurus field specification part 100, without the user 21 specifying (refer to Step 1001-1002 in drawing 10). Therefore, unlike drawing 3, there is no specification display window of the retrieval object field and the reference thesaurus field in drawing 12. The keyword interpretation part and retrieval object field and the reference thesaurus field specification part 100 are constituted with the same function as the keyword interpretation part 22 of the example 1 besides the automatic setting function of the above-mentioned retrieval object field and the reference thesaurus field. Other portions are the same as that of drawing 2 among drawing 11.

[0066]Below, the automatic setting of the keyword interpretation part and retrieval object field, the retrieval object field by the reference thesaurus field specification part 100, and the reference thesaurus field is explained. Here, since the keyword (user keyword) is specified by the user 21, a biparite field is chosen and specified based on each similarity with this user keyword, the retrieval object field, and the reference thesaurus field. The retrieval object field useful to search and the reference thesaurus field are the strong candidates of the above-mentioned retrieval object field and the biparite field which carries out automatic setting with the reference thesaurus field specification part 100. It is thought that it can divide roughly into two kinds, the field which presents the synonym connected with search in associative memory as the useful retrieval object field and a reference thesaurus field (only henceforth a field), and the field which presents a different synonym from a viewpoint or a conceptual level. The similarity between user keywords is calculated and, as for the thing of similarity which has the large former, and the latter, the small thing of similarity corresponds.

[0067]The similarity between a user keyword and a field can be found, for example with the following calculation methods. First, a field is vectorized. Only the number in many order extracts the word which appears in each field, and it normalizes. However, particles which come out frequently except. Here, suppose that five words are extracted in many order, and let these words be basic words. For example, the appearance frequency in the network field presupposes that it is e-mail 3 system 2isdn 2internet 1 cellular-phone 1.

[0068]Next, a user keyword is vectorized. In all the fields, a user keyword, for example, "Computer", assumes that the following words are adjoined by the following number of times. If this is put in order and it is a vector, scsi 4 file 2 soft 1 system 1isdn 1 will become a vector of a user keyword.

[0069]Next, the similarity between a user keyword and a field is calculated. Similarity between a user keyword and a field = it is considered as the inner product of an item which corresponded. Although a conflicting item is used for absolute value calculation of normalization of a vector, it is not used for the molecule of an inner product. The words which overlap with the user keyword between fields in the above-mentioned example are "isdn" and a "system", and it is in the network field (2, 2).

By a user keyword (1, 1)

It is ********. Therefore, similarity is set to (2x1+2x1)/(square root of square root x23 of 19) =0.19 as x=3x3+2x2+2x2+1x1+1x1=19y=4x4+2x2+1x1+1x1+1x1=23. It asks for similarity by the vector of 1000 to 5000 word actually.

[0070]If it states concretely, it will be thought that the specification of the large field of similarity with a user keyword is useful when obtaining a detailed additional keyword, and specification of the small field of similarity is useful to conversion of a viewpoint. For example, although a "terminal" means a "computer" to the user 21 of computer business, it means a "telephone" to the user 21 of a telephone industry. Since it becomes difficult to search the document from a different viewpoint so that it is a special keyword, as for a certain forge fire with a keyword with a special user keyword, the above-mentioned retrieval object field and the reference thesaurus field specification part 100 will be set up so that the small field of similarity may be chosen and specified. on the contrary, a user keyword -- **** -- when it is a general keyword, the above-mentioned retrieval object field and the reference thesaurus field specification part 100 will be set up so that the same or large field of similarity as the user keyword may be chosen and specified.

[0071]The above-mentioned retrieval object field and the reference thesaurus field specification part 100, the beginning -- the largest field of similarity -- after that and the next -- three sorts of fields with large similarity -- choose and make it specified that it says from descending of similarity, or the smallest field of similarity chooses it as the beginning or the last, and is specified as it, or various setting out is possible. The keyword interpretation part and retrieval object field and the reference thesaurus field specification part 100, A user keyword is extended using the thesaurus 23 of the field specified by the dictionary 24 and self, For example, after displaying some extended keywords on monitor display and making the user 21 choose, The retrieval part 26 is made to refer to the procedure same about retrieval object part Nouchi's retrieval object document group specified by self as the example 1, and, finally study of the dictionary 24 and the thesaurus 23 and case registration to the search case base 25 are performed.

[0072]<Effect of the example 4> Since the automatic setting of the retrieval object field and the

reference thesaurus field was made to be carried out by specifying a user keyword according to the example 4 as stated above, By setting up suitably the selection technique of each field by which automatic setting is carried out, it is effective in the ability to carry out without the search which converted the detailed search by large each field of similarity or the viewpoint by small each field of similarity applying the time and effort of the user's 21 each field specification. In addition, there is the same effect as the example 1. If it specifies sequentially from large each field of similarity with a user keyword and search is advanced, it is effective in search time shortening all the fields rather than a system conventionally which is searched at once.

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

---

[Field of the Invention]This invention relates to the electronic document search method which searches a desired document using a keyword out of two or more electronic documents (it is also only called a document in this specification.).
[0002]

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

[Description of the Prior Art]In recent years, progress of the document electronization which records on a storage or draws [ request / of paperless issue etc. ] up a document on a storage as electronic data from the start by using the contents of the paper document as electronic data is remarkable. Since the electronic document on the above-mentioned storage can also perform the search electronically, when searching the document (document which suited the user's retrieval object) desired out of a lot of documents, compared with a paper document, its usefulness is very high.
[0003]Search of an electronic document conventionally the word and the phrase (in this specification, it is only called the keyword.) used as a key, The retrieval required using what was combined with logical symbols, such as NOT which means that AND which means that two or more keywords are contained in the same document, OR which means that either of two or more keywords is contained in the same document, or a keyword is not contained in a document, is performing. Since a user's burden of keyword selection becomes large only by a keyword, there are also the existing dictionary and a thing which extends a keyword mainly using English-Japanese, a Japanese-English dictionary, or a synonym dictionary (thesaurus), and was searched. Not the binary of 1 (truth) or 0 (imitation) but the continuous value of a before [ 0-1 ] is taken to a keyword, and the fuzzy search whose search of the document containing the search word near 1 is enabled is also proposed.
[0004]conventionally the field (retrieval object field) to search is not divided into some, and it searches to one field common about any keywords, i.e., one huge field, and a thesaurus is also in the situation where it is only that there is what [ a ] is common to all the fields.

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

---

<Effect of the example 1> As opposed to the retrieval required [ according to / as stated above / the example 1 ] by specification of the retrieval object field, the reference thesaurus field, and a user keyword, Since it was made to display on monitor display with the search word and grade which used the searched document for the search, it is effective in the ability of search of only the document to desire to carry out more appropriately than the conventional method. Since the dictionary 24 and the thesaurus 23 which were used learn automatically based on the search word selected arbitrarily in the search word displayed on a user keyword, an extended keyword, and monitor display etc., Search of the document which the know how of retrieval methods, such as the user 21 keyword specification and extension, could be accumulated, and search of the document to desire improved, and contained the new word, the technical term, and the foreign language is effective in the ability to carry out now more flexibly than before and easily. The large reference thesaurus field of similarity with the retrieval object field (or the reference thesaurus field)

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

---

[Problem(s) to be Solved by the Invention]Since the new word, technical term, or foreign language which is not contained in an existing dictionary or thesaurus is not considered at all by conventional technology as mentioned above, Search of the document which cannot respond to a huge new word, a technical term, or a foreign language even if it uses the method of extending and searching a keyword, using the dictionary and thesaurus of these existing, but contains a new word etc. was impossible or remarkably difficult.
[0006]Since it would always search for [ all the ] a field even when the retrieval object field of the document to desire can predict to some extent, since there was only a thing which has the retrieval object field and a thesaurus common to all the fields, search took time.
[0007]Use of only mere English-Japanese and a Japanese-English dictionary is insufficient about extension of a keyword, therefore, pass use of the reading (Chinese character kana) dictionary of a Chinese character, or an English reading (English kana) dictionary, or proofreading [ as / in WWW or Net News ] -- use of many dictionaries, such as use etc. of the misspelling dictionary to the document which is not, can be considered. If not only the dictionary that extracts such synonymous words but the various thesauri which extract a synonym will be used, the total of the dictionary used for search or a thesaurus will increase. Then, the user needed to choose which dictionary and thesaurus are used and required time and effort for the preparation before retrieval execution. The efforts of search did not bear fruit considering time and effort, and the document to desire was not obtained.
[0008]This invention is made that the problem of the above-mentioned conventional technology should be solved.

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

[Means for Solving the Problem]The next composition is used for this invention in order to solve above-mentioned SUBJECT.
<Composition 1> by specifying two or more retrieval objects and thesauri of the request field as a seed respectively out of a retrieval object and a thesaurus by which the field division was carried out, and specifying a desired keyword, A word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus, A search word in the specification field of the above-mentioned retrieval object in an index on which a search word which carried out logical combination to the specified above-mentioned keyword, obtained an extended keyword, and was created according to each retrieval object field, and information which specifies a document containing the search word are made to come to correspond, Search document corresponding using the above-mentioned extended keyword, and a searched document, It displays on monitor display with a grade called for by a search word and the computing method set up beforehand which were used for the search, And an electronic document search method performing study of the above-mentioned dictionary and a thesaurus based on a search word selected arbitrarily in a search word displayed on the specified above-mentioned keyword, the above-mentioned extended keyword obtained from this keyword, and the above-mentioned monitor display.
[0010]<Composition 2> by specifying a retrieval object of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a thesaurus and by each above-mentioned specification of these retrieval objects, a keyword, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus to the specified above-mentioned keyword, and obtaining an extended keyword.
[0011]<Composition 3> by specifying a thesaurus of the request field and specifying a desired keyword in the electronic document search method according to claim 1, It is specified by field of a retrieval object and by each above-mentioned specification of these thesauri, a keyword,

and a retrieval object. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned thesaurus and a retrieval object to the specified above-mentioned keyword, and obtaining an extended keyword.

[0012]<Composition 4> in the electronic document search method according to claim 1, by specifying a desired keyword, It is specified by each field of a retrieval object and a thesaurus, and by each above-mentioned specification of these keywords, a retrieval object, and a thesaurus. An electronic document search method carrying out logical combination of the word selected based on each specification field and a dictionary set up beforehand of the above-mentioned retrieval object and a thesaurus to the specified above-mentioned keyword, and obtaining an extended keyword.

[0013]

[Embodiment of the Invention]Hereafter, it explains using a drawing per example of this invention.

<<Example 1>>

<The composition of the example 1 and operation> The flow chart which shows the example 1 of the electronic document search method according [ drawing 1 ] to this invention, and drawing 2 are the explanatory views of a search system in which the example 1 of this invention method was applied. As shown in drawing 2, a search system here is provided with the keyword interpretation part 22, the thesaurus 23 (23a - 23c--), the dictionary 24, the search case base 25, the retrieval part 26, and the index 27.

[0014]The keyword for search (user keyword) as which the user 21 specified the above-mentioned keyword interpretation part 22, the retrieval object field, and the field of the thesaurus 23 to refer to

---

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

---

[Brief Description of the Drawings]
[Drawing 1]It is a flow chart which shows the example 1 of this invention method.
[Drawing 2]It is the explanatory view of a search system in which the example 1 of this invention method was applied.
[Drawing 3]It is a figure showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 2.
[Drawing 4]It is a flow chart which shows the example 2 of this invention method.
[Drawing 5]It is the explanatory view of a search system in which the example 2 of this invention method was applied.
[Drawing 6]It is a figure showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 5.
[Drawing 7]It is a flow chart which shows the example 3 of this invention method.
[Drawing 8]It is the explanatory view of a search system in which the example 3 of this invention method was applied.
[Drawing 9]It is a figure showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 8.
[Drawing 10]It is a flow chart which shows the example 4 of this invention method.
[Drawing 11]It is the explanatory view of a search system in which the example 4 of this invention method was applied.
[Drawing 12]It is a figure showing an example of the monitor display display information at the time of the retrieval-required input of the search system shown in drawing 11.
[Description of Notations]
21 User
22 Keyword interpretation part
23 Thesaurus (- --)
23a Computer field thesaurus
23b Scientific discipline thesaurus

23c Social field thesaurus
24 Dictionary
25 Search case base
26 Retrieval part
27 Index
51 Keyword interpretation part and reference thesaurus field specification part
81 Keyword interpretation part and retrieval object field specification part
100 The keyword interpretation part and retrieval object field, a reference thesaurus field specification part

---

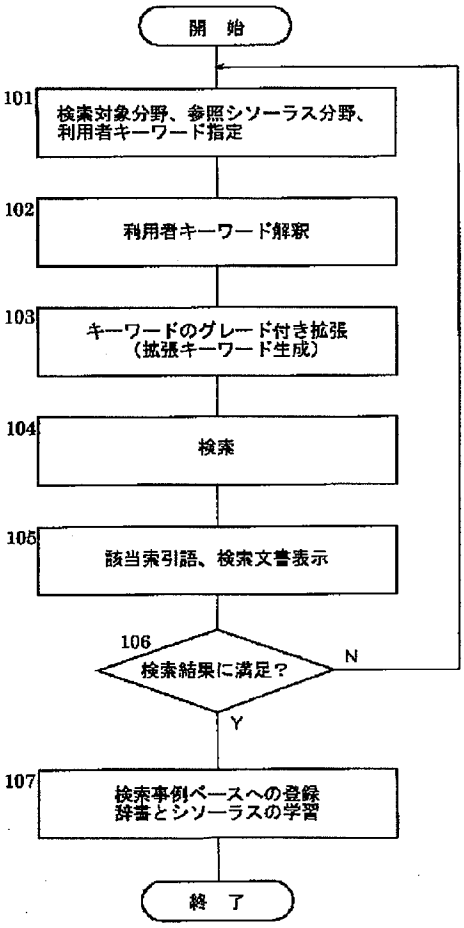
[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

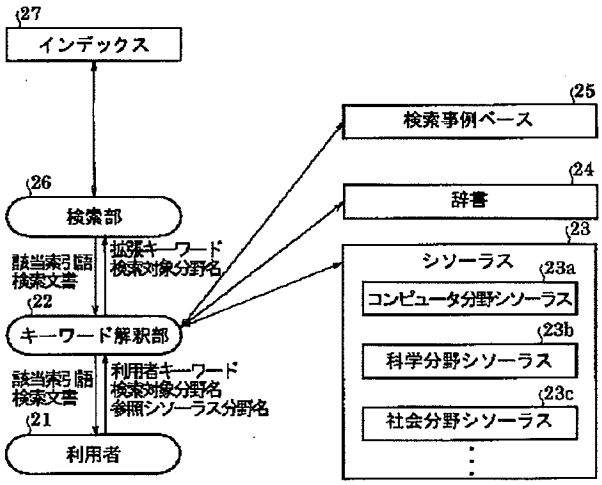
---

[Drawing 1]




本発明方法の具体例１を示すフローチャート

[Drawing 2]




本発明方法の具体例1が適用された検索システムの説明図

[Drawing 3]

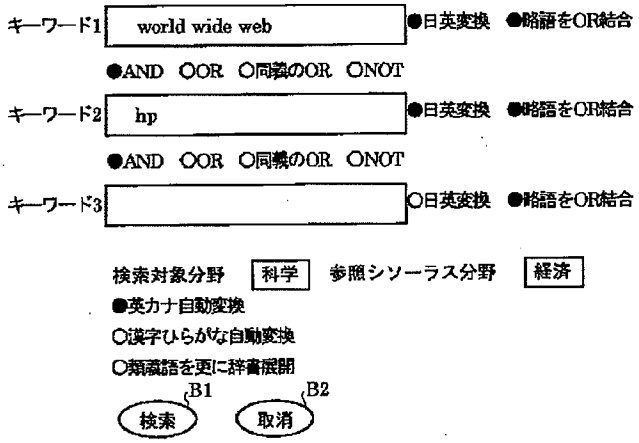
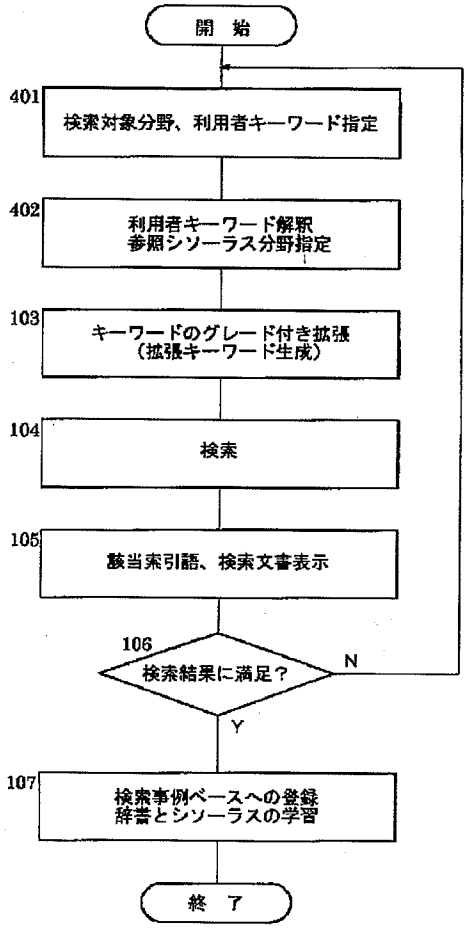


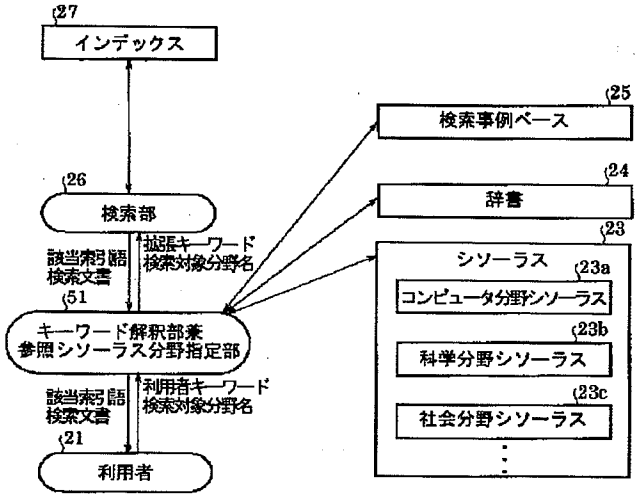

図2に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

[Drawing 4]




本発明方法の具体例２を示すフローチャート

[Drawing 5]




本発明方法の具体例２が適用された検索システムの説明図

[Drawing 6]

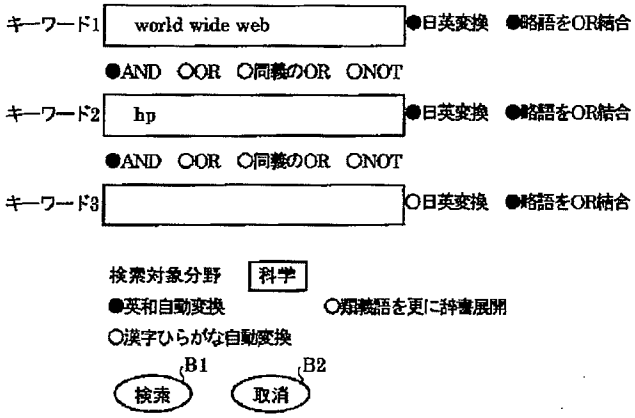
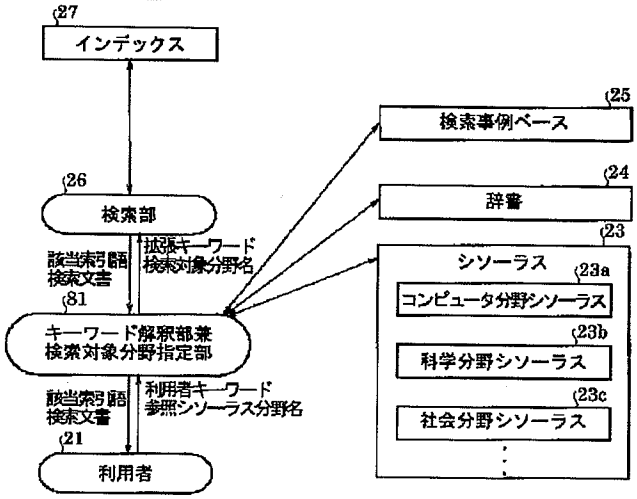


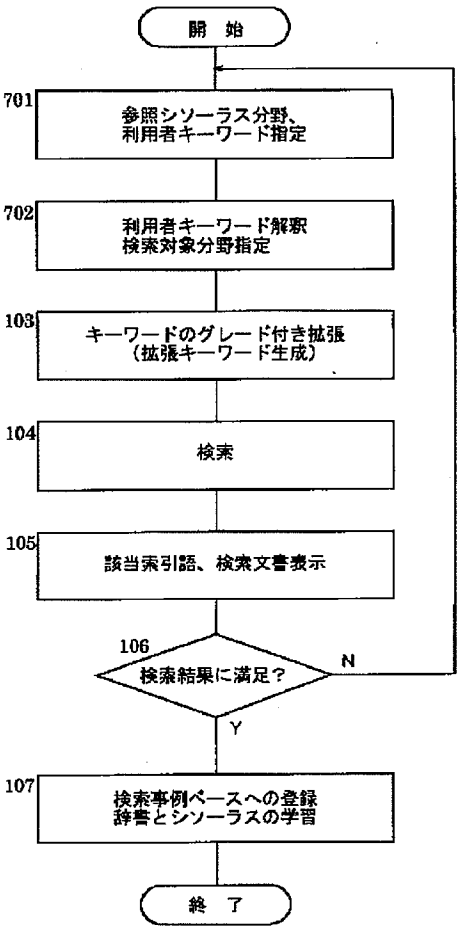

図5に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

[Drawing 8]




本発明方法の具体例3が適用された検索システムの説明図

[Drawing 7]




本発明方法の具体例３を示すフローチャート

[Drawing 9]

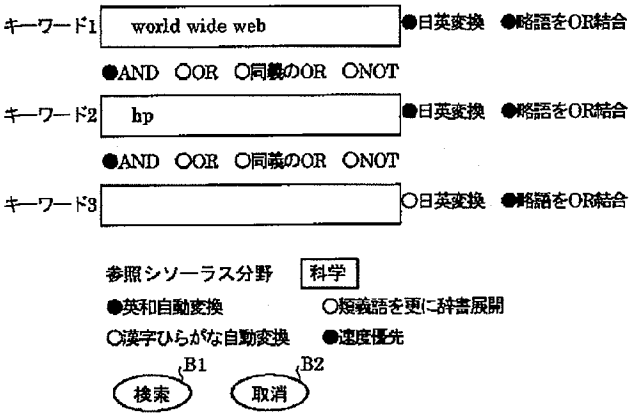
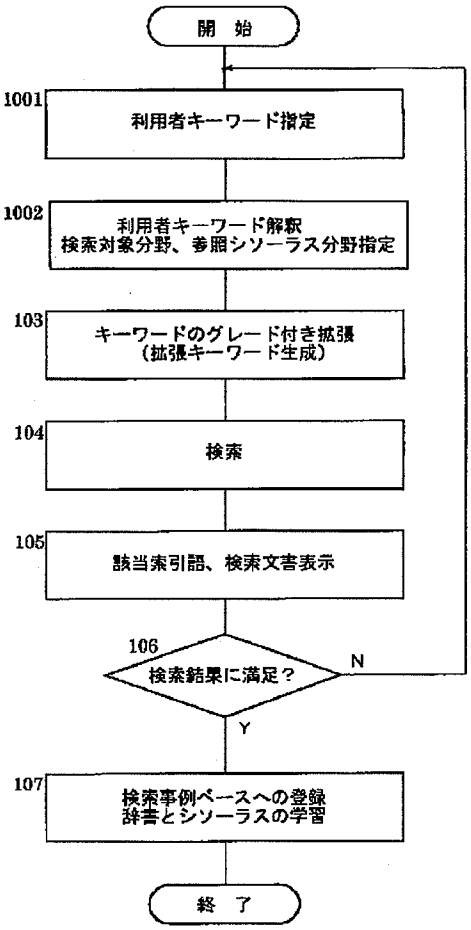


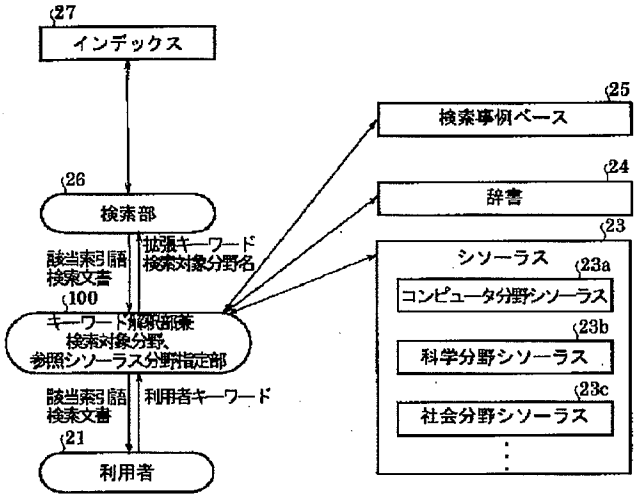

図８に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

[Drawing 10]




本発明方法の具体例４を示すフローチャート

[Drawing 11]




本発明方法の具体例４が適用された検索システムの説明図

[Drawing 12]

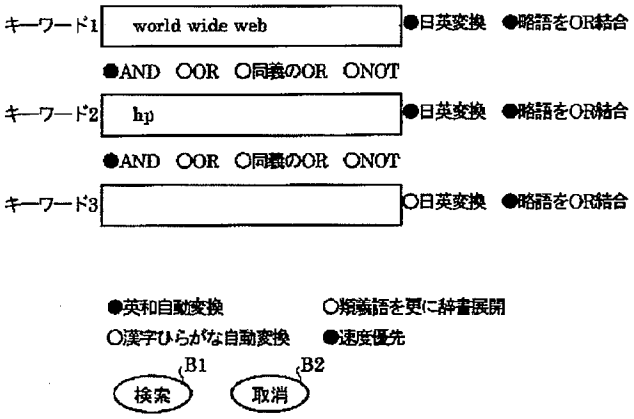



図11に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

---

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2000－331012
（P2000－331012A）
(43)公開日　平成12年11月30日(2000.11.30)

| (51)Int.Cl.[7] | 識別記号 | ＦＩ | ﾃｰﾏｺｰﾄﾞ(参考) |
|---|---|---|---|
| Ｇ０６Ｆ　17/30 | | Ｃ０６Ｆ　15/403 | ３２０Ｄ　５Ｂ０７５<br>３５０Ｃ |

審査請求　未請求　請求項の数４　ＯＬ　（全 15 頁）

(21)出願番号　特願平11－138070

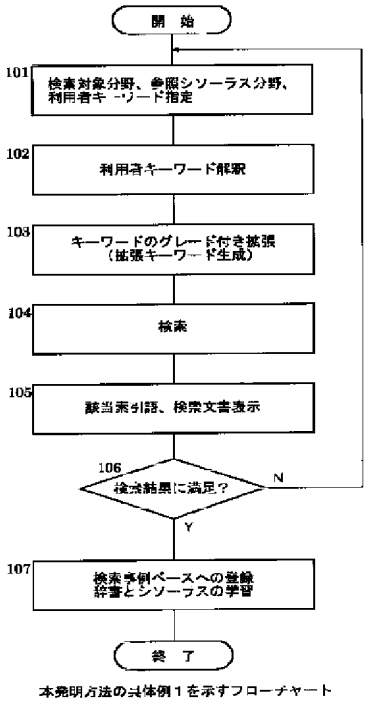
(22)出願日　平成11年５月19日(1999.5.19)

(71)出願人　000000295
沖電気工業株式会社
東京都港区虎ノ門１丁目７番12号
(72)発明者　城風　敏彦
東京都港区虎ノ門１丁目７番12号　沖電気工業株式会社内
(74)代理人　100082050
弁理士　佐藤　幸男

Ｆターム(参考)　5B075 ND02 NK02 NK35 PP12 PP13
PP22 PQ02 PQ32 PR06 QM01
QM02 QM05 QM08 QP03 QS11

(54)【発明の名称】　電子化文書検索方法

(57)【要約】
【課題】　望む文書の検索の適正化、迅速化を図る。
【解決手段】　分野分けされた検索対象，シソーラス中から所望分野の検索対象，シソーラスを指定すると共に所望のキーワードを指定することにより、検索対象，シソーラス指定分野及び辞書に基づき選定された語を、前記キーワードに論理結合して拡張キーワードを得、検索対象分野に応じて作成された索引語とそれを含む文書を特定する情報とを対応させてなるインデックス中の検索対象指定分野における索引語と、上記拡張キーワードとを用いて該当文書を検索し、その文書を、検索に用いた索引語及びグレードと共にモニタ画面に表示し、かつ辞書及びシソーラスの学習を行うこととする。




本発明方法の実施例１を示すフローチャート

【特許請求の範囲】
【請求項１】　各々複数種に分野分けされた検索対象及びシソーラスの中から所望分野の検索対象及びシソーラスを指定すると共に所望のキーワードを指定することにより、前記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された前記キーワードに論理結合して拡張キーワードを得、
各検索対象分野に応じて作成された索引語とその索引語を含む文書を特定する情報とを対応させてなるインデックス中の前記検索対象の指定分野における索引語と、前記拡張キーワードとを用いて該当文書を検索し、
検索された文書を、その検索に用いた索引語及び予め設定された算出法により求められたグレードと共にモニタ画面に表示し、
かつ、指定された前記キーワード、このキーワードから得られた前記拡張キーワード及び前記モニタ画面に表示された索引語中の任意に選択した索引語に基づいて前記辞書及びシソーラスの学習を行うことを特徴とする電子化文書検索方法。
【請求項２】　請求項１に記載の電子化文書検索方法において、所望分野の検索対象を指定すると共に所望のキーワードを指定することにより、シソーラスの分野が指定され、それら検索対象、キーワード及びシソーラスの前記各指定により、前記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された前記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。
【請求項３】　請求項１に記載の電子化文書検索方法において、所望分野のシソーラスを指定すると共に所望のキーワードを指定することにより、検索対象の分野が指定され、それらシソーラス、キーワード及び検索対象の前記各指定により、前記シソーラス、検索対象の各指定分野及び予め設定された辞書に基づき選定された語を、指定された前記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。
【請求項４】　請求項１に記載の電子化文書検索方法において、所望のキーワードを指定することにより、検索対象及びシソーラスの各分野が指定され、それらキーワード、検索対象及びシソーラスの前記各指定により、前記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された前記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は複数の電子化文書（本明細書において単に文書ともいう。）の中から所望の文書をキーワードを用いて検索する電子化文書検索方法に関するものである。
【０００２】
【従来の技術】近年、ペーパーレス化等の要請から、紙文書の内容を電子データとして記憶媒体に記録、あるいは文書を初めから電子データとして記憶媒体上に作成する文書電子化の進展が目覚ましい。上記記憶媒体上の電子化文書は、その検索も電子的に行い得るので、大量の文書中から望む文書（利用者の検索目的にかなった文書）を検索する場合に、紙文書に比べて極めて有用性が高い。
【０００３】従来、電子化文書の検索は、キーとなるワードやフレーズ（本明細書において単にキーワードという。）を、複数のキーワードが同じ文書に含まれることを意味するＡＮＤ、複数のキーワードのいずれかが同じ文書に含まれることを意味するＯＲ又はキーワードが文書に含まれないことを意味するＮＯＴ等の論理記号で結合させたものを用いた検索要求により行っている。また、キーワードのみでは利用者のキーワード選択の負担が大きくなるので、既存の辞書、主として英和，和英辞書や類義語辞書（シソーラス）を用い、キーワードを拡張して検索するようにしたものもある。更に、キーワードに対して１（真）か０（偽）かの２値ではなく、０～１までの間の連続的な値をとり、１に近い索引語を含む文書をも検索可能とするファジィ検索も提案されている。
【０００４】また従来は、検索する分野（検索対象分野）をいくつかに分けることはなく、いかなるキーワードについても共通の１分野、すなわち膨大な１つの分野に対して検索を行い、またシソーラスも全分野共通のものが１つあるだけという状況にある。
【０００５】
【発明が解決しようとする課題】上記のように従来技術では、既存の辞書やシソーラスに含まれていない新造語、専門用語あるいは外国語等には何ら配慮されていないため、それら既存の辞書やシソーラスを用い、キーワードを拡張して検索する方法を用いても膨大な新造語、専門用語あるいは外国語には対応できず、新造語等を含む文書の検索が不能又は著しく困難であった。
【０００６】また、検索対象分野やシソーラスが全分野共通のものが１つあるだけなので、望む文書の検索対象分野がある程度予測できる場合でも、常に全分野対象に検索を行うことになるので、検索に時間がかかった。
【０００７】更に、キーワードの拡張に関し、単なる英和，和英辞書のみの使用では不充分であり、したがって漢字の読み（漢字かな）辞書や英語の読み（英語カナ）辞書の使用、あるいはWWWやネットニュースにおけるような校正を経ていない文書に対するミススペル辞書の使用等、多数の辞書の使用が考えられる。このような同義語を抽出する辞書だけでなく、類義語を抽出する各種シソーラスをも使うことになると、検索に使用する辞書やシソーラスの総数が増大する。そこで利用者は、どの辞書やシソーラスを使用するかを選択する必要があり、

検索実行前の準備に手間がかかった。また、手間の割には検索の成果が上がらず、望む文書が得られなかった。

【０００８】本発明は、上記従来技術の問題を解決すべくなされたものである。

【０００９】

【課題を解決するための手段】本発明は、上述課題を解決するため次の構成を採用する。

〈構成１〉各々複数種に分野分けされた検索対象及びシソーラスの中から所望分野の検索対象及びシソーラスを指定すると共に所望のキーワードを指定することにより、上記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された上記キーワードに論理結合して拡張キーワードを得、各検索対象分野に応じて作成された索引語とその索引語を含む文書を特定する情報とを対応させてなるインデックス中の上記検索対象の指定分野における索引語と、上記拡張キーワードとを用いて該当文書を検索し、検索された文書を、その検索に用いた索引語及び予め設定された算出法により求められたグレードと共にモニタ画面に表示し、かつ、指定された上記キーワード、このキーワードから得られた上記拡張キーワード及び上記モニタ画面に表示された索引語中の任意に選択した索引語に基づいて上記辞書及びシソーラスの学習を行うことを特徴とする電子化文書検索方法。

【００１０】〈構成２〉請求項１に記載の電子化文書検索方法において、所望分野の検索対象を指定すると共に所望のキーワードを指定することにより、シソーラスの分野が指定され、それら検索対象、キーワード及びシソーラスの上記各指定により、上記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された上記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。

【００１１】〈構成３〉請求項１に記載の電子化文書検索方法において、所望分野のシソーラスを指定すると共に所望のキーワードを指定することにより、検索対象の分野が指定され、それらシソーラス、キーワード及び検索対象の上記各指定により、上記シソーラス、検索対象の各指定分野及び予め設定された辞書に基づき選定された語を、指定された上記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。

【００１２】〈構成４〉請求項１に記載の電子化文書検索方法において、所望のキーワードを指定することにより、検索対象及びシソーラスの各分野が指定され、それらキーワード、検索対象及びシソーラスの上記各指定により、上記検索対象、シソーラスの各指定分野及び予め設定された辞書に基づき選定された語を、指定された上記キーワードに論理結合して拡張キーワードを得ることを特徴とする電子化文書検索方法。

【００１３】

【発明の実施の形態】以下、本発明の具体例につき図面を用いて説明する。

《具体例１》

〈具体例１の構成，動作〉図１は本発明による電子化文書検索方法の具体例１を示すフローチャート、図２は本発明方法の具体例１が適用された検索システムの説明図である。図２に示すように、ここでの検索システムは、キーワード解釈部２２、シソーラス２３（２３ａ～２３ｃ…）、辞書２４、検索事例ベース２５、検索部２６及びインデックス２７を備えてなる。

【００１４】上記キーワード解釈部２２は、利用者２１が指定した検索用キーワード（利用者キーワード）、検索対象分野及び参照するシソーラス２３の分野（参照シソーラス分野）からなる検索要求を受け付けて、拡張キーワード及び検索対象分野を検索部２６に与える構成部である。ここで利用者キーワードは、通常、複数のキーワードが論理結合され、また部分一致の記号（ワイルドカード）を含んでなる。またキーワード解釈部２２は、指定された利用者キーワード、ここでは部分一致の記号（ワイルドカード）を含んだ利用者キーワードを解釈し、部分一致の種類（完全一致、前方一致等）を判別する機能及びワイルドカードを切り離した純粋なキーワードを抽出する機能をもつ。更にキーワード解釈部２２は、抽出されたキーワードの組が検索事例ベース２５にそのまま存在するか否かを確認し、存在すればその事例における拡張後のキーワード（拡張キーワード）の組を抽出し、それをモニタ画面（図示せず）を介して利用者２１に提示し、利用者２１の必要に応じた修正を待って検索部２６に与える機能をもつ。上記キーワードの組が検索事例ベース２５にそのまま存在しなければ、利用者２１によって指定されたシソーラス２３あるいは辞書２４にそれらのキーワードが見出し語として登録されているか否かを各々確認し、登録されていればその見出し語から得られる語（同義語，類義語）と上記組をなすキーワードの各々とＯＲ結合された拡張キーワードとして検索部２６に与える。

【００１５】上記シソーラス２３は、ここではコンピュータ分野シソーラス２３ａ、科学分野シソーラス２３ｂ及び社会分野シソーラス２３ｃ等のように予め検索対象分野毎に作成されており、また見出し語に対する類義語の類似度付きのファジィシソーラスとなっている。上記辞書２４は、ここでは英和、和英、漢字かな、英語カナ、ミススペル及び略語の６つの辞書を備えてなる。

【００１６】上記検索事例ベース２５は、過去の検索例（検索事例）を、１又は複数、ここでは複数のキーワード、検索対象分野及び参照シソーラス分野の組で記憶しており、それらのうち２つまでが利用者２１によって指定されたら残りの１つを自動的に指定する機能をもつもので、検索要求の補完に使用される。例えば、検索事

例ベース２５に、
キーワード：ホームページ　作成
検索対象分野：ネットワーク
参照シソーラス分野：科学
という過去の検索例があった場合、
キーワード：ホームページ　作成
検索対象分野：ネットワーク
まで検索要求が指定されると、
参照シソーラス分野：科学
を補完してモニタ画面に表示し、利用者２１の確認がとれると参照シソーラス分野として“科学”を自動指定する。

【００１７】検索部２６は、キーワード解釈部２２からの拡張キーワードと検索対象分野とを受け、インデックス２７を参照して該当文書を検索する構成部である。すなわち検索部２６は、検索対象文書群中の各文書から抽出された索引語群で構成されたインデックス２７中の、上記キーワード解釈部２２から与えられた検索対象分野内の検索対象文書群に属する索引語群中の索引語の各々と上記拡張キーワードとを比較し、一致する索引語を含む文書をグレード付きで抽出してモニタ画面に表示することで、検索結果を利用者２１に与えるものである。上記グレードは、例えば利用者キーワードそのものの論理結合を満たす文書については１とし、類義語を含む文書は１以下、隣接条件を満たす文書は１以上、キーワードを直接含まなければ１以下とする。検索によりキーワードと一致した索引語は検索文書と共にモニタ画面に表示されるが、そのうち所望の索引語を利用者２１がモニタ画面上で選ぶと、それが類義語であればシソーラス２３中の上記検索対象分野のシソーラスに、同義語であれば辞書２４に各々登録（学習）される。指定されたキーワード、検索対象分野及び参照シソーラス分野の組は検索事例ベース２５に検索事例として登録される。これらの学習，登録処理は上記キーワード解釈部２２によって行われる。

【００１８】上記インデックス２７は、ここでは次のように作成されている。インデックス２７の作成に当たっては、そのサイズを小さくして検索を速めるため、漢字、平仮名、片仮名、英字、数字等の字種が異なる単語の重複がないようになされ、字種の区切りで索引語の区切りとされる。例えば、通常の索引語切出しツールを使用すると“情報フィルタリング”からは、“情報フィルタリング”、“情報”及び“フィルタリング”の３つを抽出することになるが、これでは索引語数が増加してインデックス２７のサイズが増大し、検索に時間がかかるようになる。そのためここでは、“情報”及び“フィルタリング”の２つだけを索引語とする。“情報フィルタリング”という検索要求に対しては、検索要求時に“情報”及び“フィルタリング”が１語隣にあるという指定（隣接のＡＮＤ＝ＮＡＮＤ指定）をして検索することで補う。インデックス２７は、例えば索引語をそのままコード順にソートしたものと、“計算機”を“機算計”のように逆順にしてコード順にソートしたものとの２つが作成，使用される。後者のものは後方一致検索に使用されるが、必須のものではない。インデックス２７における索引語から文書名を検索するためのインデックス構造は公知のデータベースにおける検索に用いられるものと同様である。

【００１９】上述検索システムの検索要求入力時におけるモニタ画面表示内容の一例を図３に示す。この図において、白丸“○”及び黒丸“●”は各々ポインティングデバイス（図示せず）で操作される機能選択用釦スイッチで、白丸はオフ、黒丸はオンを表す。ここでは、第１キーワードに“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”が、第２キーワードに“ｈｐ”が、ＡＮＤなる論理結合をもって入力されている。検索対象分野は“科学”が、参照シソーラス分野は“経済”が各々入力されている。上記ポインティングデバイスで検索釦スイッチＢ１を操作すれば検索が開始され、取消釦スイッチＢ２を操作すれば全ての入力操作、動作中においてはその動作が取り消され、キーワード、分野、釦スイッチは初期状態（空白あるいはオフ）に戻される。

【００２０】以下、本発明方法の具体例１を図１，図３を併用して述べる。ステップ１０１では、利用者２１によりキーワード（利用者キーワード）、検索対象分野及び参照シソーラス分野が指定（入力）される。図３を例に採って述べると、利用者キーワードは、「略語をＯＲ結合する」ための機能選択用釦スイッチがオン（“●”）になっていることから分かるように、第１の利用者キーワード（キーワード１）“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”について、その略語“ｗｗｗ”がＯＲ結合により拡張されるよう指定されている。

【００２１】またここでは、１フィールド内に複数語並べて書かれた利用者キーワードは、各語が隣接のＡＮＤで結合されていると解釈され、通常のＡＮＤ結合の場合はフィールドを変更して入力することとされている。したがって、図３に例示する第１の利用者キーワード（キーワード１）“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”は、基本的にはこの３つの単語がある同じ文書に存在し、かつ、この順序で連続して出現する文書を検索せよという検索要求と解釈されるが、同時に“ｗｏｒｌｄ”と“ｗｉｄｅ”、“ｗｉｄｅ”と“ｗｅｂ”は各々隣接のＡＮＤ結合であると解釈される。このような論理結合による検索は通常の読みの語順に従って実行される。第２の利用者キーワード（キーワード２）“ｈｐ”は、“ｗｗｗ”がＯＲ結合されて拡張された上記利用者キーワード“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”とＡＮＤ結合されるものと解釈される。

【００２２】このような利用者キーワードの解釈，拡張は後述ステップ１０２，１０３で行われることになる

が、こうした解釈，拡張がなされることを前提として図３に示すモニタ画面上で利用者キーワードが指定される。検索事例ベース２５による、過去の検索例に基づくキーワード、検索対象分野又は参照シソーラス分野の前述自動指定も行われる。

【００２３】図３には例示されていないが、利用者キーワードは、通常、部分一致等を指定するためのいくつかの記号が付されて入力される。例えば、（ａ）任意の文字列と一致する記号（ワイルドカード）として“＊”、（ｂ）１つのキーワード中に２つ以上の単語が“―”で結ばれているか、隣接していることを示す記号として“＿”、（ｃ）１つのキーワード中に２つの単語が両端に存在するか、１つ目の単語は前方一致、２つめの単語は後方一致で、これら２つが隣接していることを示す記号として“＋”、が設定され、適宜キーワードに付される。

【００２４】上記（ａ）によれば、ＵＮＩＸのシェルの正規表現と同じく、“＊”によって前方一致、後方一致、中間一致、両端一致等、柔軟な指定ができる。（ｂ）によれば、“―”で結ばれた英語の熟語（ａｎｇｌｅｄ－ｓｈｏｔ等）、（ｃ）によれば、助詞で結ばれた日本語の熟語（情報処理の資格試験＝情報処理資格試験等）について、各々有効な検索を行える。例えば、単語“情報”と“試験”をもつキーワードを、情報＊試験ＯＲ（情報＊ＮＡＮＤ＊試験）とする。これによって“情報処理試験”、“情報の資格試験”、“情報処理の資格試験”、情報処理資格試験”等が同時に検索できることになる。

【００２５】またここでは、利用者キーワードの論理結合を２段まで許すこととし、１段目の論理結合はＡＮＤ、隣接のＡＮＤ及びＯＲの３種類、２段目の論理結合はＡＮＤ、ＯＲ、略語のＯＲ及びＮＯＴの４種類とする。例えば、ＮＡＮＤを論理記号のＮＡＮＤではなく隣接のＡＮＤ、ＡＯＲを略語のＯＲとして、

（ｗｏｒｌｄ ＮＡＮＤ ｗｉｄｅ ＮＡＮＤ ｗｅｂ）ＡＮＤ（ｈｏｍｅ ｐａｇｅ＊）ＡＮＤ日弁連

のようになる。ＡＯＲは、辞書２４中の略語辞書への登録（後述ステップ１０７参照）を主目的とする場合に用いられるもので、例えば、

（ｗｏｒｌｄ ＮＡＮＤ ｗｉｄｅ ＮＡＮＤ ｗｅｂ）ＡＯＲ ｗｗｗ

とキーワード指定すれば、“ｗｏｒｌｄ ｗｉｄｅ ｗｅｂ”の略語として“ｗｗｗ”が略語辞書に必ず登録される。なおここでは、“ｗｗｗ”に他の正式名称（語）があってもＯＲ結合しないこととされている。

【００２６】ステップ１０２では、利用者キーワードが解釈される。このステップ１０２では、上述利用者キーワードの解釈に加え、指定された利用者キーワードから一致の種類と純粋なキーワードの抽出が行われる。例として、利用者キーワードがｉｎｆｏ＊ｆｉｌｔｅｒであれば、

一致の種類：両端一致

キーワード：（ｉｎｆｏ，ｆｉｌｔｅｒ）

と解釈して検索システム中のメモリに保存される。

【００２７】ステップ１０３では、利用者キーワードが、辞書２４及び指定されたシソーラス２３を用いて拡張され、拡張キーワードとして検索部２６に与えられる。具体的には、利用者キーワードが辞書２４中の各見出し語と順次比較対照され、一致した見出し語から得られる語とＯＲ結合により拡張（ＯＲ拡張）され、拡張キーワードとして検索部２６に与えられる。

【００２８】辞書２４の種類によっては、分野や文脈によって意味が変わらないもの、例えば英和，和英辞書等と、変わるものとがあるので、キーワード解釈部２２は利用者２１が指定した分野のシソーラス２３に基づきどの意味をとるかを選択する。例えば、利用者キーワード“ｍｏｎｋｅｙ”を“猿”と拡張することに問題はないであろうから、この場合は両者をＯＲ結合して、すなわち“ｍｏｎｋｅｙ”ＯＲ“猿”を拡張キーワードとして検索部２６に与える。利用者キーワードが“ｃｏｍｐ”であったとすると、これは情報の分野において正式な単語“ｃｏｍｐｕｔｅｒ”の略語であるから両者をＯＲ結合して、すなわち“ｃｏｍｐ”ＯＲ“ｃｏｍｐｕｔｅｒ”とＯＲ拡張し、これを拡張キーワードとして検索部２６に与える。“ｃｏｍｐ＊”と前方一致で指定すると“ｃｏｍｐａｒｅ”等、関係のない語を含む文書が検索されてしまうので、ここではこのような拡張は行わない。

【００２９】複合語の略語は曖昧性が高いので、例えば次のように拡張する。すなわち、利用者キーワードが“ｈｐ”であったとすると、その正式な語（複合語）“ｈｏｍｅ ｐａｇｅ”とＮＡＮＤ（隣接のＡＮＤ）結合して、“ｈｏｍｅ”ＮＡＮＤ“ｐａｇｅ”と拡張する。キーワード“ｈｐ”には“ｈｏｍｅ ｐａｒｔｙ”という意味もある。この場合、キーワード解釈部２２は、利用者２１が指定した検索対象分野が娯楽や生活であれば“ｈｏｍｅ ｐａｒｔｙ”であると、コンピュータやネットワークであれば“ｈｏｍｅ ｐａｇｅ”であると判別し、上記と同様にＮＡＮＤ結合して拡張する。

【００３０】ＯＲ結合される単語にはグレードが付される。このグレードは計算により求められるが、その根拠となるものは、ここでは単語の連接確率である。「連接」とは、単語同士、例えば単語ｉと単語ｊが接近して出現することを意味し、単語ｉと単語ｊが、前後何単語以内に出現、１文書内に出現（共起）あるいは特定文書集合内に出現（共起）というように種々の態様が考えられる。連接確率は、適宜の態様、ここでは隣接（前後１語で出現）なる態様が選択されて下式（１）で求められ、上記グレードとして用いられる。

連接確率Ｗｉｊ＝（単語ｉと単語ｊが連接した回数）／（単語ｉと単語ｊのどちらかが出現した回数）　…（１）

このような連接確率（グレード）は分野毎に求められる。

【００３１】あるキーワードｋｉとある単語ｋｊとの連接確率Ｗｉｊは、例えばキーワード“メール”に対して、

“電子”０.５

“ネットワーク”０,３

“受信”０.２

のように表される。

【００３２】出現する単語全てについて連接確率を計算し保存しておくことは、そのために必要とするメモリ容量が多大になることや、得られた連接確率が実用上、どの程度信頼のおけるものとなるかを考慮すると、必ずしも得策とはいい難い。したがって実際には、ほぼ同じ意味の複数の単語をグループ化し、そのうちの１つの単語（代表単語）について連接確率を計算し、その値をグループに属する全ての単語の連接確率とされる。例えば、“ネットワーク”、“Ｎｅｔｗｏｒｋ”及び“電網”等で“ネットワークグループ”を形成し、そのうち上記“ネットワーク”について求められた連接確率を“ネットワークグループ”内の他の単語、すなわち“Ｎｅｔｗｏｒｋ”及び“電網”等の連接確率としても用いる。これら“ネットワーク”、“Ｎｅｔｗｏｒｋ”及び“電網”等は、ここでは同義語として辞書２４中に、連接確率（グレード）と共に登録されている。

【００３３】いま、利用者キーワードとして“コンピュータ”が指定されたものとし、また、辞書２４には、英語対日本語の関係で、Ｃｏｍｐｕｔｅｒ，計算機が、英語対カタカナの関係で、Ｃｏｍｐｕｔｅｒ，コンピュータが、日本語対ひらがなの関係で、計算機，けいさんきが、英語略語としてＣｏｍｐｕｔｅｒ，Ｃｏｍｐ．が、日本語略語として電子計算機，計算機が、表記のゆらぎとしてコンピュータ，コンピューターが、各々登録されていたとする。一方、上述したように“ネットワーク”及び“Ｎｅｔｗｏｒｋ”は同義語として辞書２４中に登録されているので、“Ｃｏｍｐｕｔｅｒ　Ｎｅｔｗｏｒｋ”も“計算機ネットワーク”も同じものとしてその連接確率（グレード）を計算できる。これにより、“Ｃｏｍｐｕｔｅｒ　Ｎｅｔｗｏｒｋ”も“計算機ネットワーク”も同じグレードで検索されることになる。すなわち、“ＣｏｍｐｕｔｅｒＮｅｔｗｏｒｋ”が検索されれば“計算機ネットワーク”も検索されることになり、また、後述するように検索結果（文書）にグレードが付される場合には“Ｃｏｍｐｕｔｅｒ　Ｎｅｔｗｏｒｋ”を含む文書と“計算機ネットワーク”を含む文書とは同じ値のグレードが付されることになる。

【００３４】次に、利用者キーワード“ｈｐ”について説明する。“ｈｐ”は略語であり、正式な語（複合語）として“ｈｏｍｅ　ｐａｇｅ”と“ｈｏｍｅ　ｐａｒｔｙ”の２つがあるものとする。いま、コンピュータ分野シソーラス２３ａにおいて、“ｈｏｍｅ　ｐａｇｅ”として３００回、“ｈｏｍｅ　ｐａｒｔｙ”として１００回出現したとすると、その場合の各語のコンピュータ分野でのグレードは、例えばグレード＝（求める語の出現回数）／（いずれかの語の最大出現回数）とすると、

“ｈｏｍｅ　ｐａｇｅ”：３００／３００＝１

“ｈｏｍｅ　ｐａｒｔｙ”：１００／３００＝０．３３３…

となる。したがって、コンピュータ分野シソーラス２３ａを指定することにより、“ｈｏｍｅ　ｐａｇｅ”を含む文書は、“ｈｏｍｅ　ｐａｒｔｙ”を含む文書より必ず大きなグレードで検索されることになる。また、検索結果（文書）にグレードが付される場合には、“ｈｏｍｅ　ｐａｇｅ”を含む文書と“ｈｏｍｅ　ｐａｒｔｙ”を含む文書の上記グレードの大小に応じた比率でグレードが付されることになる。

【００３５】上述例は単語が隣接した複合語（熟語）の場合を述べたもので、この場合には検索対象分野と一致する分野のシソーラス２３を用いた方が望む文書の検索上、有効とされるが、検索対象分野とは異なる分野のシソーラス２３を用いた方がよい場合もある。パソコン（パーソナルコンピュータ）が機械であることはコンピュータ分野では自明であり、したがって、利用者キーワード“パソコン”から、それと検索対象分野が一致するコンピュータ分野のシソーラス２３を用いて類義語“機械”をＯＲ拡張することことは困難であると考えられるからである。このような場合は、視点を変えるために、例えばシソーラス２３中の経済分野シソーラスや娯楽分野シソーラス（いずれも図示せず）というような検索対象分野とは異なる分野のシソーラス２３を用いて類義語をＯＲ拡張する。“パソコン”と“機械”のように概念に上下関係がある単語同士の場合には、検索対象分野とは異なる分野のシソーラス２３を用いて類義語をＯＲ拡張した方が検索結果が向上することが多い点からも、このようなＯＲ拡張が有効であるいえる。本具体例１において、検索要求に当たり、参照するシソーラス分野を任意に選択可能（ステップ１０１参照）としているのは、そのためである。“パーソナルコンピュータ”と“ワークステーション”のような概念の上下関係がない、同レベルの単語における類義語のＯＲ拡張に当たっては、検索対象分野と参照シソーラス分野とを一致させるという基本手法が守られる。

【００３６】なお、上式（１）で求まる連接確率Ｗｉｊはある単語と他の単語との類似度をも表す。したがって、シソーラス２３における類義語の類似度計算にも連接確率Ｗｉｊの計算式である上式（１）が適用できる。

【００３７】ステップ１０４では、文書検索が行われる。具体的には、キーワード解釈部２６からの拡張キーワードがインデックス２７中の各索引語と順次比較対照され、一致した索引語を含む文書の抽出が行われる。抽出された文書には、拡張キーワードがもつグレードがその文書のグレードとして付される。検索には公知の探索法、例えば２分探索法が用いられる。

【００３８】いま、拡張キーワードに“ｉｎｆｏ＊”が含まれているとすると、これは“ｉｎｆｏ”の前方一致検索であり、この場合、例えば、

ｉｎｆｏ
ｉｎｆｏ．
ｉｎｆｏｒｍ
ｉｎｆｏｍａｔｉｏｎ
ｉｎｆｏｒｍａｔｉｏｎ

が該当する索引語となり、

ｉｎｆｏ　５００文書
ｉｎｆｏ．２００文書
ｉｎｆｏｒｍ　３００文書
ｉｎｆｏｍａｔｉｏｎ　１００文書
ｉｎｆｏｒｍａｔｉｏｎ　１０５６文書

等という抽出結果（該当索引語に対する文書抽出数）が得られる。

【００３９】そして、各索引語を含む文書に対し、拡張キーワード中の他のキーワードの論理結合によって更に計算（ＯＲで和集合、ＡＮＤで積集合等）し、望む文書群を絞込み抽出する。ＡＮＤは算術積か最小値、ＯＲは算術和か最大値、ＮＯＴは差で計算する。隣接のＡＮＤの場合は、まずそれを通常の（隣接していない）ＡＮＤであると仮定して検索を行い、次にこれにより抽出された文書群中において、ＡＮＤで結ばれた２つのキーワードが隣接しているか否かをチェックし、隣接している文書を抽出して検索結果とする。この検索結果はステップ１０５でモニタ画面に表示されるが、通常のＡＮＤであると仮定して検索し、抽出された文書群を検索結果として表示するようにしてもよい。この際、隣接している文書のグレードを上げておけば、隣接していない文書との区別が容易になる。ここでは、隣接している文書のグレードを１．１倍してあり、検索の段階に応じた検索結果の表示，確認に便宜が計られている。

【００４０】ステップ１０５では、検索結果の表示が行われる。すなわち検索が終了すると、それにより抽出された文書（検索文書）がモニタ画面に表示される。検索結果である検索文書がいかなる索引語で検索されたものかを知らせるため、検索文書はその検索に用いられた索引語との対応で表示される。また、各検索文書はそのグレードが付されて表示される。ここでのグレードは、各検索文書の抽出過程において計算された上記各グレード（値）を加算あるいは乗算等、ここでは乗算することで求められた値とされている。

【００４１】検索文書の表示は、まず各索引語に対する検索文書数の表示、次に特定の索引語をポインティングデバイス（図示せず）により指示することによるその索引語で抽出された文書名の表示、続いて特定の文書名を上記ポインティングデバイスで指示することによるその文書の該当ページ（索引語が記述されているページ）の表示等、種々の段階表示が可能である。なお検索文書の表示としては、最終的にその文書を特定できる情報が表示されればよく、例えばそれが書籍であれば書籍の題名、著者、発行年月日、発行所等が、雑誌であればそれら題名等に加えてシリアル番号が、論文であればそれが載った学会誌名、論文のタイトル、発表者、発行年月日、発行所等が該当する。

【００４２】ステップ１０６では、検索結果（検索文書）の確認が行われる。すなわちこのステップ１０６では、検索結果に満足したか否かの判定がなされる。ステップ１０５で検索結果が表示され、望む文書が得られたときには利用者２１の検索結果に満足する旨の操作によりステップ１０７に処理が移る。望む文書が得られない等、利用者２１が検索結果に満足しないときには、満足するまでステップ１０１～１０６が繰り返される。ここでは、利用者２１が最初に指定したキーワード（利用者キーワード）、検索対象分野名及び参照シソーラス分野名は検索システムのバッファメモリに残すことになっているので、検索繰返し時における利用者２１の操作としては、通常、キーワードを追加指定したり上記各分野名を変更指定するといった微調整で済む。

【００４３】ステップ１０７では、辞書２４及びシソーラス２３の学習、検索事例ベース２５への事例登録が行われる。具体的には、利用者キーワードにＯＲ結合された略語は結合前の利用者キーワードの略語として辞書２４中の略語辞書に登録される。例えば、利用者キーワード“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”を“ｗｗｗ”という略語をＯＲ結合して検索を行ったところ、望む文書が抽出された場合（ステップ１０６において結果満足と判定された場合）は、“ｗｗｗ”が“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”の略語として辞書２４中の略語辞書に登録される。なお図３は、“ｗｗｗ”が“ｗｏｒｌｄ　ｗｉｄｅ　ｗｅｂ”の略語として辞書２４中の略語辞書に初めから登録されており、機能選択用釦スイッチのオンでその略語“ｗｗｗ”がＯＲ結合されるようなされた場合を例示したもので、ここでの例とは異なる。

【００４４】検索によりキーワードと一致した索引語は、ステップ１０５においてモニタ画面に表示されるが、そのうち所望の索引語を利用者２１がモニタ画面上で選ぶと、それが類義語であればシソーラス２３中の上記検索対象分野のシソーラスに、同義語であれば辞書２４に各々登録（学習）される。

【００４５】利用者２１が指定したシソーラス２３も、利用者２１の指定したキーワード（利用者キーワード）

の内容によって以下のように学習される。すなわち、検索システムの稼働後は、当該検索システムに対して多くの利用者２１…から多くの利用者キーワードが与えられるが、この際、ＮＡＮＤ（隣接のＡＮＤ）結合指定のキーワード相互は類似度が高いとして連接確率を大きくする。例えば、利用者キーワードが“マルチ”ＮＡＮＤ“メディア”であった場合、新たな連接確率Ｗｉｊ＾を、Ｋｉ＝マルチ、Ｋｊ＝メディアとして、

Ｗｉｊ＾＝ｗｉｊ＋｛（ＫｉとＫｊがＡＮＤ結合した回数）／（ＫｉとＫｊのどちらかが出現した回数）｝×（１－Ｗｉｊ）　…式（４）

とする。いま、キーワード“マルチ”に対してキーワード“メディア”の連接確率Ｗｉｊ＾が０．７５であったとすると、そのシソーラス２３には、“マルチ”又は“メディア”に対する“メディア”又は“マルチ”の類似度が０．７５であるとして登録（学習）される。これにより、よく連接するキーワード間の類似度が大きくなり、その後の検索時に、“マルチ”又は“メディア”のいずれか一方が利用者キーワードとなったり、利用者キーワードに含まれたりした場合に、他方の語“メディア”又は“マルチ”を含む文書のグレードが大きくなり、その文書の望む文書としての検索がしやすくなる。

【００４６】指定されたキーワード、検索対象分野及び参照シソーラス分野の組は検索事例ベース２５に検索事例として登録される。これらの学習，登録処理は上記キーワード解釈部２２によって行われる。

【００４７】〈具体例１の効果〉以上述べたように具体例１によれば、検索対象分野、参照シソーラス分野及び利用者キーワードの指定による検索要求に対して、検索された文書をその検索に用いた索引語及びグレードと共にモニタ画面に表示するようにしたので、望む文書だけの検索が従来方法よりも適切に行い得るという効果がある。また、利用者キーワード、拡張キーワード及びモニタ画面に表示された索引語中の任意に選択した索引語等に基づいて、使用した辞書２４やシソーラス２３が自動的に学習するので、利用者２１のキーワード指定，拡張等の検索手法のノウハウが蓄積でき、望む文書の検索が向上し、また、新造語、専門用語、外国語を含んだ文書の検索が従来より柔軟かつ容易に行えるようになるという効果もある。検索対象分野（又は参照シソーラス分野）との類似度の大きい参照シソーラス分野（又は検索対象分野）から順に指定して検索を進めれば、全分野を一度に検索する従来システムよりも検索時間が短縮するという効果もある。

【００４８】《具体例２》

〈具体例２の構成，動作〉図４は本発明による電子化文書検索方法の具体例２を示すフローチャート、図５は本発明方法の具体例２が適用された検索システムの説明図、図６は図５に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。なお、これら図４～図６において、図１～図３と同一又は相当部分には同一符号を付してその説明を省略する。図５に示すように、ここでの検索システムは、キーワード解釈部兼参照シソーラス分野指定部５１、シソーラス２３（２３ａ～２３ｃ…）、辞書２４、検索事例ベース２５、検索部２６及びインデックス２７を備えてなる。

【００４９】上記キーワード解釈部兼参照シソーラス分野指定部５１は、利用者２１が指定した利用者キーワード及び検索対象分野からなる検索要求を受け付けて、参照シソーラス分野を選択，指定し、拡張キーワード及び検索対象分野を検索部２６に与える構成部である。すなわち具体例２では、参照シソーラス分野は利用者２１が指定することなく、キーワード解釈部兼参照シソーラス分野指定部５１にて自動指定されるものである（図４中のステップ４０１，４０２参照）。したがって図６には、図３と異なり、検索対象分野の指定表示窓の右隣に参照シソーラス分野の指定表示窓がない。キーワード解釈部兼参照シソーラス分野指定部５１は上記参照シソーラス分野の自動指定機能の他、具体例１のキーワード解釈部２２と同様の機能をもって構成されている。図５中、その他の部分は図２と同様である。

【００５０】以下に、キーワード解釈部兼参照シソーラス分野指定部５１による参照シソーラス分野の自動指定について説明する。まず、検索対象分野は利用者２１により指定されているので、その検索対象分野と同一分野のシソーラスは上記参照シソーラス分野指定部５１で自動指定するシソーラスの第１候補である。また、検索に有用な分野のシソーラスも有力な候補として挙げられる。検索に有用な分野のシソーラスとしては、連想記憶的に連接する類義語を提示してくれる分野のシソーラスと、異なった視点あるいは概念レベルからの類義語を提示してくれる分野のシソーラスの２種類に大別できると考えられる。検索対象分野との間の類似度を計算して、前者は類似度の大きいもの、後者は類似度の小さいものが該当する。

【００５１】分野間の類似度は、例えば以下のような計算方法により求まる。まず、分野のベクトル化を行う。各々の分野に出現する単語を多い順にある数だけ抽出し、正規化する。ただし、頻繁に出てくる助詞等は除外する。ここでは多い順に５単語を抽出することとし、これらの単語を基本単語とする。例えばネットワーク分野での出現回数が、

ｅ－ｍａｉｌ　３
システム　２
ｉｓｄｎ　２
ｉｎｔｅｒｎｅｔ　１
携帯電話　１

であり、コンピュータ分野のベクトルは、

ｓｃｓｉ　４

ファイル　２
ソフト　１
システム　１
ｉｓｄｎ　１
であるとする。
【００５２】次に、分野間の類似度を計算する。２つの分野間の類似度＝２つの正規化ベクトルの一致した項目の内積とする。一致しない項目はベクトルの正規化の絶対値計算には使うが、内積の分子には用いない。上述例では、両分野において重複している単語は“ｉｓｄｎ”と“システム”であり、
ネットワーク分野では（２，２）
コンピュータ分野では（１，１）
の組合せである。したがって類似度は、
ｘ＝３×３＋２×２＋２×２＋１×１＋１×１＝１９
ｙ＝４×４＋２×２＋１×１＋１×１＋１×１＝２３
として、
（２×１＋２×１）／（１９の平方根×２３の平方根）
＝０．１９
となる。なお、実際には１０００～５０００単語のベクトルで類似度を求める。
【００５３】具体的に述べると、検索対象分野との類似度の大きい分野のシソーラスの使用は詳細な追加キーワードを得るときに有用であり、類似度の小さい分野のシソーラスの使用は視点の転換に有用と考えられる。例えば“端末”はコンピュータ業界の利用者２１には“コンピュータ”を意味するが、電話業界の利用者２１には“電話”を意味する。専門的な分野であるほど異なった視点からの文書は検索し難くなるので、利用者２１により指定された検索対象分野が専門的な分野であればあるほど、類似度の小さい分野のシソーラスが選択，指定されるように上記参照シソーラス分野指定部５１が設定されることになる。逆に、利用者２１により指定された検索対象分野が極く一般的な分野であるときには、その検索対象分野と同一の又は類似度の大きい分野のシソーラスが選択，指定されるように上記参照シソーラス分野指定部５１が設定されることになる。
【００５４】キーワード解釈部兼参照シソーラス分野指定部５１は、辞書２４及び自身が指定した分野のシソーラス２３を用いて利用者キーワードの拡張を行い、例えばいくつかの拡張キーワードをモニタ画面に表示して利用者２１に選択させた後、利用者２１が指定した検索対象分野内の検索対象文書群について具体例１と同様の手順で検索部２６に検索させ、最後に、辞書２４及びシソーラス２３の学習、検索事例ベース２５への事例登録を行う。
【００５５】〈具体例２の効果〉以上述べたように具体例２によれば、検索対象分野及び利用者キーワードを指定することにより参照シソーラス分野が自動指定されるようにしたので、自動指定される参照シソーラス分野の選択手法を適宜設定することにより、類似度の大きい分野のシソーラスによる詳細な検索あるいは類似度の小さい分野のシソーラスによる視点を転換した検索が利用者２１の参照シソーラス分野指定の手間をかけずに行えるという効果がある。その他、具体例１と同様な効果がある。検索対象分野との類似度の大きい参照シソーラス分野から順に指定して検索を進めれば、全分野を一度に検索する従来システムよりも検索時間が短縮するという効果もある。
【００５６】《具体例３》
〈具体例３の構成，動作〉図７は本発明による電子化文書検索方法の具体例３を示すフローチャート、図８は本発明方法の具体例３が適用された検索システムの説明図、図９は図８に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。なお、これら図７～図９において、図１～図３と同一又は相当部分には同一符号を付してその説明を省略する。図８に示すように、ここでの検索システムは、キーワード解釈部兼検索対象分野指定部８１、シソーラス２３（２３ａ～２３ｃ…）、辞書２４、検索事例ベース２５、検索部２６及びインデックス２７を備えてなる。
【００５７】上記キーワード解釈部兼検索対象分野指定部８１は、利用者２１が指定した利用者キーワード及び参照シソーラス分野からなる検索要求を受け付けて、検索対象分野を選択，指定し、拡張キーワード及び検索対象分野を検索部２６に与える構成部である。すなわち具体例３では、検索対象分野は利用者２１が指定することなく、キーワード解釈部兼検索対象分野指定部８１にて自動指定されるものである（図７中のステップ７０１，７０２参照）。したがって図９には、図３と異なり検索対象分野の指定表示窓がない。キーワード解釈部兼検索対象分野指定部８１は上記検索対象分野の自動指定機能の他、具体例１のキーワード解釈部２２と同様の機能をもって構成されている。図８中、その他の部分は図２と同様である。
【００５８】以下に、キーワード解釈部兼検索対象分野指定部８１による検索対象分野の自動指定について説明する。まず、参照シソーラス分野は利用者２１により指定されているので、その参照シソーラス分野と同一の検索対象分野は上記検索対象分野指定部８１で自動指定する検索対象分野の第１候補である。また、検索に有用な検索対象分野も有力な候補として挙げられる。検索に有用な検索対象分野としては、連想記憶的に連接する類義語を提示してくれる検索対象分野と、異なった視点あるいは概念レベルからの類義語を提示してくれる検索対象分野の２種類に大別できると考えられる。参照シソーラス分野との間の類似度を計算して、前者は類似度の大きいもの、後者は類似度の小さいものが該当する。
【００５９】分野間の類似度は、例えば以下のような計算方法により求まる。まず、分野のベクトル化を行う。

各々の分野に出現する単語を多い順にある数だけ抽出し、正規化する。ただし、頻繁に出てくる助詞等は除外する。ここでは多い順に５単語を抽出することとし、これらの単語を基本単語とする。例えばネットワーク分野での出現回数が、
ｅ－ｍａｉｌ　３
システム　２
ｉｓｄｎ　２
ｉｎｔｅｒｎｅｔ　１
携帯電話　１
であり、コンピュータ分野のベクトルは、
ｓｃｓｉ　４
ファイル　２
ソフト　１
システム　１
ｉｓｄｎ　１
であるとする。

【００６０】次に、分野間の類似度を計算する。２つの分野間の類似度＝２つの正規化ベクトルの一致した項目の内積とする。一致しない項目はベクトルの正規化の絶対値計算には使うが、内積の分子には用いない。上述例では、両分野において重複している単語は“ｉｓｄｎ”と“システム”であり、
ネットワーク分野では（２，２）
コンピュータ分野では（１，１）
の組合せである。したがって類似度は、
ｘ＝３×３＋２×２＋２×２＋１×１＋１×１＝１９
ｙ＝４×４＋２×２＋１×１＋１×１＋１×１＝２３
として、
（２×１＋２×１）／（１９の平方根×２３の平方根）
＝０．１９
となる。なお、実際には１０００～５０００単語のベクトルで類似度を求める。

【００６１】具体的に述べると、参照シソーラス分野との類似度の大きい検索対象分野の指定は詳細な追加キーワードを得るときに有用であり、類似度の小さい検索対象分野の指定は視点の転換に有用と考えられる。例えば“端末”はコンピュータ業界の利用者２１には“コンピュータ”を意味するが、電話業界の利用者２１には“電話”を意味する。専門的な分野であるほど異なった視点からの文書は検索し難くなるので、利用者２１により指定された参照シソーラス分野が専門的な分野であればあるほど、類似度の小さい検索対象分野が選択，指定されるように上記検索対象分野指定部８１が設定されることになる。逆に、利用者２１により指定された参照シソーラス分野が極く一般的な分野であるときには、その参照シソーラス分野と同一の又は類似度の大きい検索対象分野が選択，指定されるように上記検索対象分野指定部８１が設定されることになる。

【００６２】また上記検索対象分野指定部８１は、最初に類似度の最も大きい検索対象分野が、その後、次に類似度の大きい３種の検索対象分野が、というように類似度の大きい順から選択，指定されるようにしたり、あるいは最初又は最後に類似度の最も小さい検索対象分野が選択，指定されるようにしたり、種々の設定が可能である。更に、キーワード解釈部兼検索対象分野指定部８１は、辞書２４及び利用者２１が指定した分野のシソーラス２３を用いて利用者キーワードの拡張を行い、例えばいくつかの拡張キーワードをモニタ画面に表示して利用者２１に選択させた後、自身が指定した検索対象分野内の検索対象文書群について具体例１と同様の手順で検索部２６に検索させ、最後に、辞書２４及びシソーラス２３の学習、検索事例ベース２５への事例登録を行う。

【００６３】〈具体例３の効果〉以上述べたように具体例３によれば、参照シソーラス分野及び利用者キーワードを指定することにより検索対象分野が自動指定されるようにしたので、自動指定される検索対象分野の選択手法を適宜設定することにより、類似度の大きい検索対象分野による詳細な検索あるいは類似度の小さい検索対象分野による視点を転換した検索が利用者２１の検索対象分野指定の手間をかけずに行えるという効果がある。その他、具体例１と同様な効果がある。参照シソーラス分野との類似度の大きい検索対象分野から順に指定して検索を進めれば、全分野を一度に検索する従来システムよりも検索時間が短縮するという効果もある。

【００６４】《具体例４》
〈具体例４の構成，動作〉図１０は本発明による電子化文書検索方法の具体例４を示すフローチャート、図１１は本発明方法の具体例４が適用された検索システムの説明図、図１２は図１１に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。なお、これら図１０～図１２において、図１～図３と同一又は相当部分には同一符号を付してその説明を省略する。図１１に示すように、ここでの検索システムは、キーワード解釈部兼検索対象分野，参照シソーラス分野指定部１００、シソーラス２３（２３ａ～２３ｃ…）、辞書２４、検索事例ベース２５、検索部２６及びインデックス２７を備えてなる。

【００６５】上記キーワード解釈部兼検索対象分野，参照シソーラス分野指定部１００は、利用者２１が指定した利用者キーワードからなる検索要求を受け付けて、検索対象分野及び参照シソーラス分野を選択，指定し、拡張キーワード及び検索対象分野を検索部２６に与える構成部である。すなわち具体例４では、検索対象分野及び参照シソーラス分野は利用者２１が指定することなく、キーワード解釈部兼検索対象分野，参照シソーラス分野指定部１００にて自動指定されるものである（図１０中のステップ１００１，１００２参照）。したがって図１２には、図３と異なり検索対象分野及び参照シソーラス分野の指定表示窓がない。キーワード解釈部兼検索対象

分野，参照シソーラス分野指定部１００は上記検索対象分野及び参照シソーラス分野の自動指定機能の他、具体例１のキーワード解釈部２２と同様の機能をもって構成されている。図１１中、その他の部分は図２と同様である。

【００６６】以下に、キーワード解釈部兼検索対象分野，参照シソーラス分野指定部１００による検索対象分野及び参照シソーラス分野の自動指定について説明する。ここでは、キーワード（利用者キーワード）が利用者２１により指定されているので、この利用者キーワードと検索対象分野及び参照シソーラス分野との各類似度をもとに両分野を選択，指定する。検索に有用な検索対象分野，参照シソーラス分野は上記検索対象分野，参照シソーラス分野指定部１００で自動指定する両分野の有力な候補である。検索に有用な検索対象分野，参照シソーラス分野（以下、単に分野という。）としては、連想記憶的に連接する類義語を提示してくれる分野と、異なった視点あるいは概念レベルからの類義語を提示してくれる分野の２種類に大別できると考えられる。利用者キーワードとの間の類似度を計算して、前者は類似度の大きいもの、後者は類似度の小さいものが該当する。

【００６７】利用者キーワードと分野間の類似度は、例えば以下のような計算方法により求まる。まず、分野のベクトル化を行う。各々の分野に出現する単語を多い順にある数だけ抽出し、正規化する。ただし、頻繁に出てくる助詞等は除外する。ここでは多い順に５単語を抽出することとし、これらの単語を基本単語とする。例えばネットワーク分野での出現回数が、

ｅ－ｍａｉｌ　３
システム　２
ｉｓｄｎ　２
ｉｎｔｅｒｎｅｔ　１
携帯電話　１

であるとする。

【００６８】次に利用者キーワードのベクトル化を行う。全分野において利用者キーワード、例えば“Ｃｏｍｐｕｔｅｒ”が、以下のような単語と以下のような回数で隣接しているものとする。これを並べてベクトルとすると、

ｓｃｓｉ　４
ファイル　２
ソフト　１
システム　１
ｉｓｄｎ　１

が、利用者キーワードのベクトルとなる。

【００６９】次に、利用者キーワードと分野間の類似度を計算する。利用者キーワードと分野間の類似度＝２つの正規化ベクトルの一致した項目の内積とする。一致しない項目はベクトルの正規化の絶対値計算には使うが、内積の分子には用いない。上述例では、利用者キーワードと分野間において重複している単語は“ｉｓｄｎ”と“システム”であり、

ネットワーク分野では（２，２）
利用者キーワードでは（１，１）

の組合せである。したがって類似度は、

ｘ＝３×３＋２×２＋２×２＋１×１＋１×１＝１９
ｙ＝４×４＋２×２＋１×１＋１×１＋１×１＝２３

として、

（２×１＋２×１）／（１９の平方根×２３の平方根）＝０．１９

となる。なお、実際には１０００～５０００単語のベクトルで類似度を求める。

【００７０】具体的に述べると、利用者キーワードとの類似度の大きい分野の指定は詳細な追加キーワードを得るときに有用であり、類似度の小さい分野の指定は視点の転換に有用と考えられる。例えば“端末”はコンピュータ業界の利用者２１には“コンピュータ”を意味するが、電話業界の利用者２１には“電話”を意味する。専門的なキーワードであるほど異なった視点からの文書は検索し難くなるので、利用者キーワードが専門的なキーワードあればあるほど、類似度の小さい分野が選択，指定されるように上記検索対象分野，参照シソーラス分野指定部１００が設定されることになる。逆に、利用者キーワードが極く一般的なキーワードであるときには、その利用者キーワードと同一の又は類似度の大きい分野が選択，指定されるように上記検索対象分野，参照シソーラス分野指定部１００が設定されることになる。

【００７１】また上記検索対象分野，参照シソーラス分野指定部１００は、最初に類似度の最も大きい分野が、その後、次に類似度の大きい３種の分野が、というように類似度の大きい順から選択，指定されるようにしたり、あるいは最初又は最後に類似度の最も小さい分野が選択，指定されるようにしたり、種々の設定が可能である。キーワード解釈部兼検索対象分野，参照シソーラス分野指定部１００は、辞書２４及び自身が指定した分野のシソーラス２３を用いて利用者キーワードの拡張を行い、例えばいくつかの拡張キーワードをモニタ画面に表示して利用者２１に選択させた後、自身が指定した検索対象分野内の検索対象文書群について具体例１と同様の手順で検索部２６に検索させ、最後に、辞書２４及びシソーラス２３の学習、検索事例ベース２５への事例登録を行う。

【００７２】〈具体例４の効果〉以上述べたように具体例４によれば、利用者キーワードを指定することにより検索対象分野及び参照シソーラス分野が自動指定されるようにしたので、自動指定される各分野の選択手法を適宜設定することにより、類似度の大きい各分野による詳細な検索あるいは類似度の小さい各分野による視点を転換した検索が利用者２１の各分野指定の手間をかけずに行えるという効果がある。その他、具体例１と同様な効

果がある。利用者キーワードとの類似度の大きい各分野から順に指定して検索を進めれば、全分野を一度に検索する従来システムよりも検索時間が短縮するという効果もある。

【図面の簡単な説明】

【図１】本発明方法の具体例１を示すフローチャートである。

【図２】本発明方法の具体例１が適用された検索システムの説明図である。

【図３】図２に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。

【図４】本発明方法の具体例２を示すフローチャートである。

【図５】本発明方法の具体例２が適用された検索システムの説明図である。

【図６】図５に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。

【図７】本発明方法の具体例３を示すフローチャートである。

【図８】本発明方法の具体例３が適用された検索システムの説明図である。

【図９】図８に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。

【図１０】本発明方法の具体例４を示すフローチャートである。

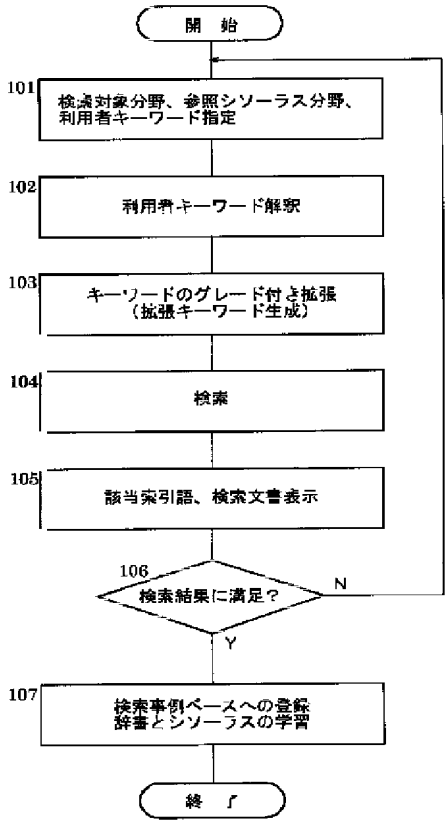
【図１１】本発明方法の具体例４が適用された検索システムの説明図である。

【図１２】図１１に示した検索システムの検索要求入力時におけるモニタ画面表示内容の一例を示す図である。

【符号の説明】

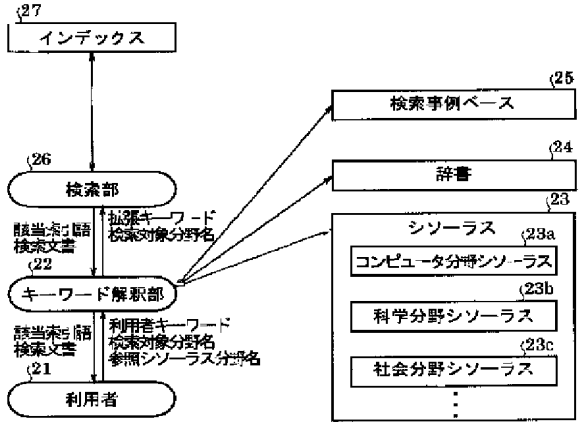
２１　利用者
２２　キーワード解釈部
２３　シソーラス（～…）
２３ａ　コンピュータ分野シソーラス
２３ｂ　科学分野シソーラス
２３ｃ　社会分野シソーラス
２４　辞書
２５　検索事例ベース
２６　検索部
２７　インデックス
５１　キーワード解釈部兼参照シソーラス分野指定部
８１　キーワード解釈部兼検索対象分野指定部
１００　キーワード解釈部兼検索対象分野，参照シソーラス分野指定部

【図１】




本発明方法の具体例１を示すフローチャート

【図２】




本発明方法の具体例１が適用された検索システムの説明図

【図3】

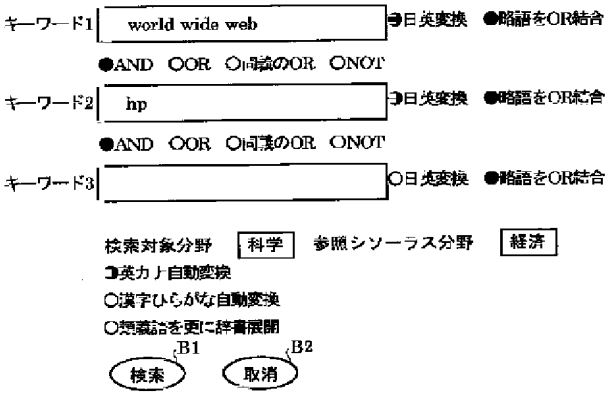
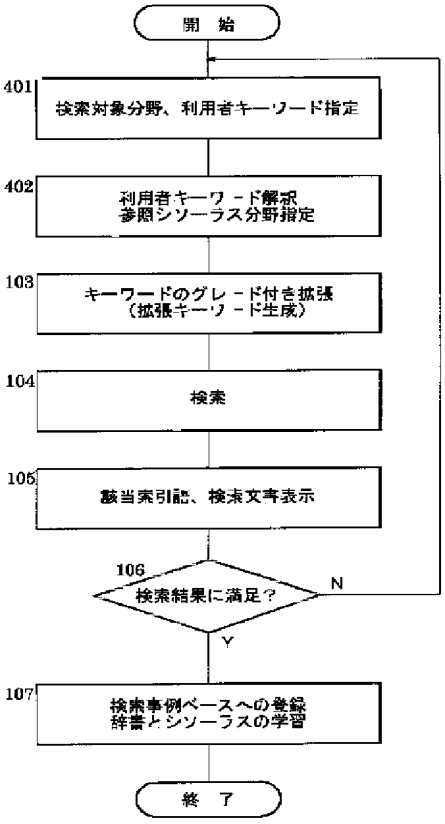


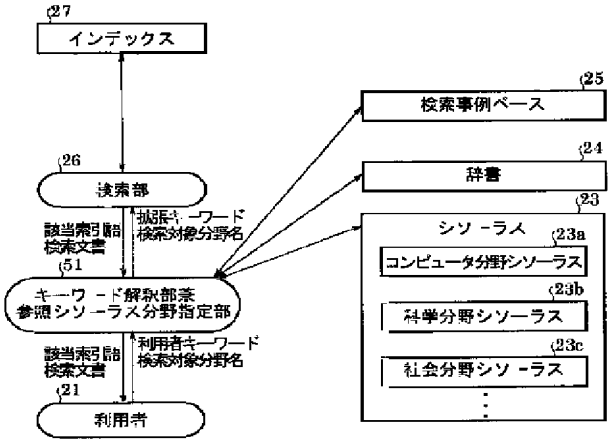

図2に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

【図4】




本発明方法の具体例2を示すフローチャート

【図5】




本発明方法の具体例2が適用された検索システムの説明図

【図6】

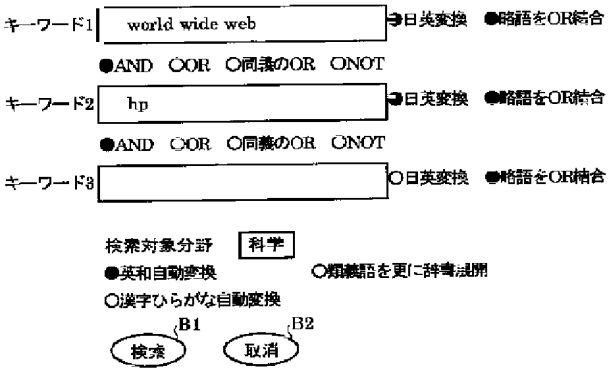
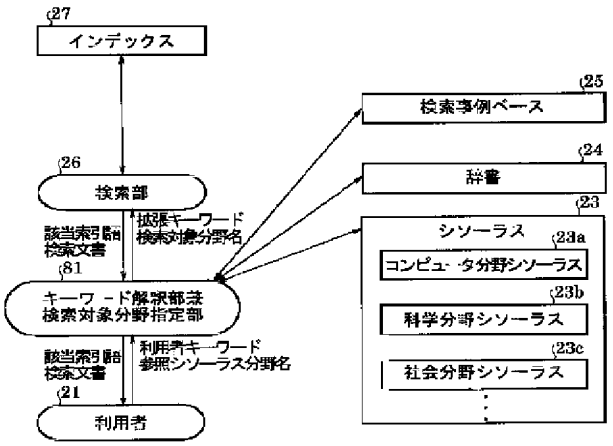


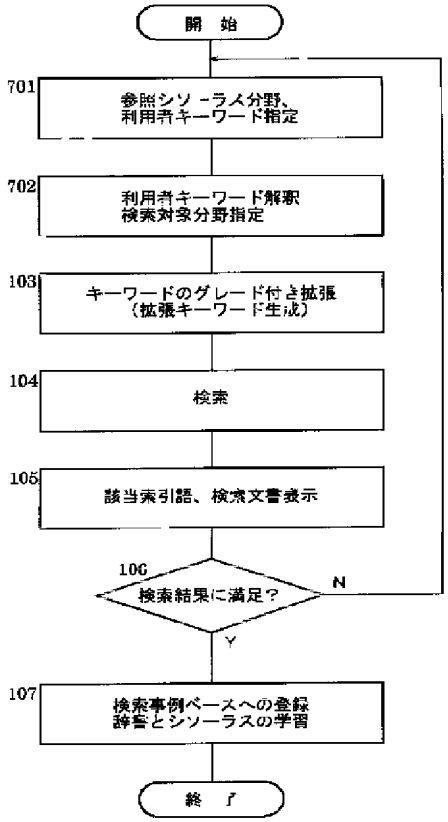

図6に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

【図8】




本発明方法の具体例3が適用された検索システムの説明図

【図７】




本発明方法の具体例３を示すフローチャート

【図９】

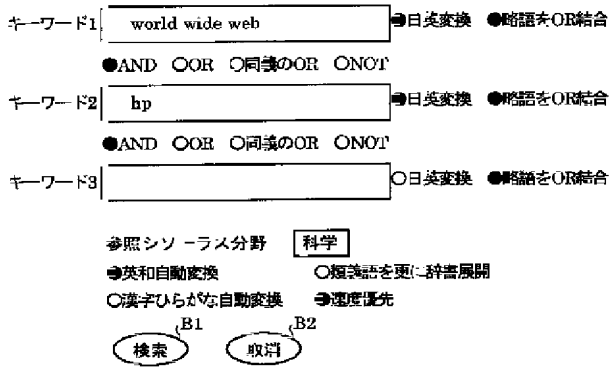
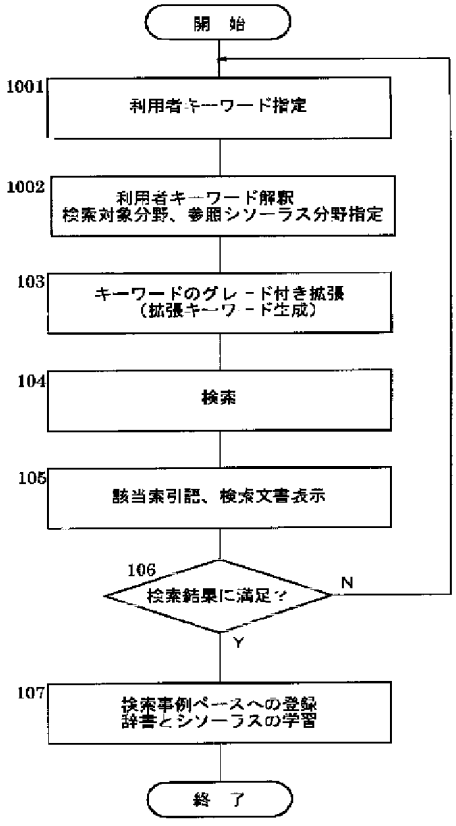


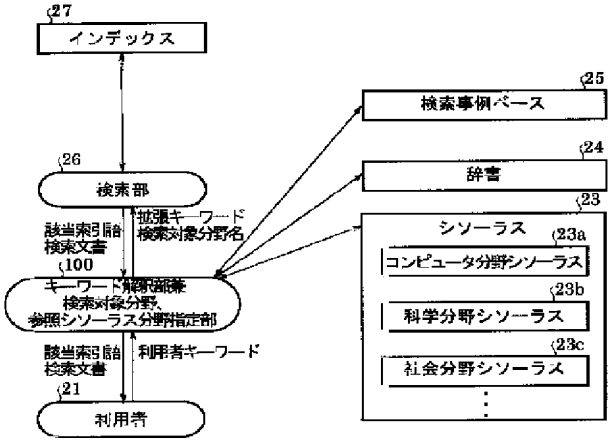

図８に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図

【図１０】




本発明方法の具体例４を示すフローチャート

【図１１】




本発明方法の具体例４が適用された検索システムの説明図

【図１２】

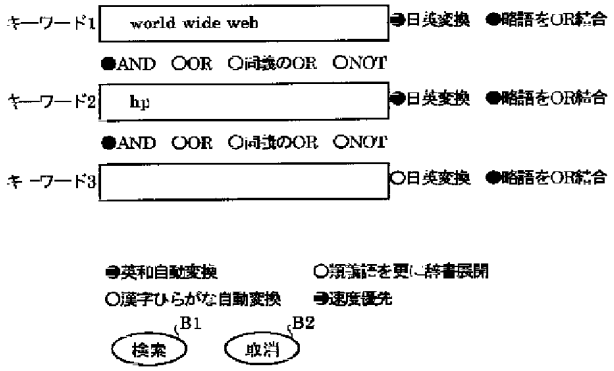



図11に示したシステムの検索要求入力時のモニタ画面表示内容例を示す図